A Computational and Experimental Study of HER2-Signaling Effects on Cellular Migration and Proliferation

by

Neil Kumar

B.S., Stanford University (2001)
M.S., Stanford University (2002)

Submitted to the Department of Chemical Engineering
in partial fulfillment of the requirements for the degree

Doctor of Philosophy in Chemical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
November 2006
[February 2007]

Signature of Author.....
Department of Chemical Engineering

Certified by.....
Douglas A. Lauffenburger
Professor of Chemical Engineering, Biological Engineering, and Biology
Thesis Supervisor

Accepted by.....
William Deen
Chairperson, Department Committee for Graduate Students

A Computational and Experimental Study of HER2-Signaling Effects on Cellular Migration and Proliferation

by

Neil Kumar

Submitted to the Department of Chemical Engineering
on November 2006, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

The fundamental question posed in this thesis is: how does a cell 'decide' to behave in a particular way? The human body is comprised of ~10^{14} cells that interpret extracellular information and respond with such behavior as migration, proliferation, apoptosis, or differentiation. Thirty years of research in the related fields of biochemistry, molecular biology, and genetics have demonstrated that, in most cases, the cellular decision-making process cannot be described or predicted by regulation of only one gene or one protein alone. Instead, it has become clear that cellular behavior is a function of information flow through multiple intracellular molecules. Furthermore, the molecules responsible for the control of cell behavior comprise a surprisingly short list, indicating that factors such as signaling dynamics and intensity coupled with combinatorial control are essential to produce the wide array of observed cell behavior. The identification of protein kinases as transducers of large amounts of intracellular information led us to pose the hypothesis that the quantitative regulation of key kinases governs cellular behavior. The goal of this thesis was to identify rules governing multi-kinase behavioral control and to then, on the basis of these rules, predict changes in cell function in response to changes in receptor expression, ligand treatment, and pharmacological intervention.

A human mammary epithelial cell (HMEC) system with varying levels of the human epidermal growth factor receptor 2 (HER2) was chosen to explore cell decision processes. HER2 overexpression is found in 30% of breast cancers and correlates with poor prognosis and increased metastasis. In particular, we investigated the effects of HER2 overexpression on signaling networks and resultant cell proliferation and migration in the presence of epidermal growth factor (EGF) or heregulin (HRG), two EGFR-family ligands that promote HER2 heterodimerization.

To investigate HER2-mediated signaling and cell behavior we developed and applied high-throughput experimental techniques to measure kinase activity and phosphorylation as well as cell proliferation and migration. Measurement of ~100 different kinases downstream of HER2 resulted in the identification of network signaling mechanisms. Application of a novel high-throughput migration

assay enabled the identification of HER2-mediated increases in cell migration due to increases in the directional persistence of movement.

Linear mapping techniques related to partial least squares regression (PLSR) defined and predicted cell behavior in response to HER2 overexpression. Combining quantitative datasets of both biological signals and behavior using PLSR, we identified subsets of kinase phosphorylation events that most critically regulate HER2-mediated migration and proliferation. Importantly, we demonstrated that our models provide predictive ability through *a priori* predictions of cell behavior in HER2-overexpressing cells. Application of linear models in response to pharmacological inhibition resulted in the *a priori* prediction of cell migration, and identified an EGFR kinase inhibitor Gefitinib as a potent inhibitor of HER2-mediated migration.

In conclusion, the application of computational linear modeling to quantitative biological signaling and behavior datasets captured systems-level regulation of cell behavior and, based on this, predicted cell migration and proliferation in response to HER2 overexpression and pharmacological inhibition. Further application of quantitative measurement together with linear modeling should enable the identification of salient cell signal-cell response elements to understand how cells make decisions and to predict how those decisions can be therapeutically manipulated.

Thesis Supervisor: Douglas A. Lauffenburger
Title: Professor of Chemical Engineering, Biological Engineering, and Biology

Acknowledgements

This thesis was made possible by the help of many fellow colleagues. Alejandro Wolf-Yadlin and Hyung-Do Kim were my two closest experimental collaborators, with whom the mass spectrometry and cell migration experiments were performed. James Evans contributed his time and effort in helping me set up the high-throughput migration assay detailed in Chapter 2. I was also fortunate to work with a number of talented undergraduates, namely Raffi Afeyan, Sonal Patel, and Sarah Sheppard. Raffi contributed to the signaling work detailed in Chapters 6 and 7, and Sarah contributed her many talents to the work presented in Chapter 7. A collaboration with AstraZeneca and the many good people in the Signaling Pathways group also helped shape the migration and signaling work in this thesis. In addition, collaboration with Bart Hendriks and David DeGraaf at Pfizer resulted in many of the ideas introduced in Chapter 8. Throughout the development of my thesis, I was very fortunate to have been guided and mentored by Muhammad Zaman, Kevin Janes, and Forest White. Muhammad, a scientist with interminable energy, helped me to focus my migration studies and provided invaluable mentorship along the way. Forest White not only lent his considerable skill to our collaborative mass spectrometry research project, but also became my 'silent advisor,' helping me to further develop my scientific thought process and providing a brand of mentorship that pushed me to a better scientist and person. Finally, Kevin Janes was an invaluable mentor and friend to me throughput the course of my PhD. His work set the basis for the studies presented herein, and the importance of his mentoring and guidance cannot be overstated.

I thank my parents and sister for all of their love and support, without which I could not have finished this PhD. In addition, many thanks to my friends and especially to those in the Lauffenburger and Griffith Laboratories that have made coming to work such fun over the past four years. In conclusion, I must thank three mentors that have defined my educational experience. Temba Maqubela, my high school chemistry teacher, deserves many thanks for, along with my father, being the reason that I became excited about studying science. Steven Chu, my undergraduate research advisor, was an extraordinary mentor who gave me confidence in my abilities as a critical thinker and scientist. Finally, I am indebted to Douglas Lauffenburger, my graduate research advisor. It is very difficult to capture the many ways Doug has contributed to this work. Obviously, the science presented herein is a product of his support and insight. But Doug's contribution goes far beyond science. It was his vision that lead me to MIT to pursue my graduate studies, and once here, his unconditional support and generosity have lead to my growth as a person and as an investigator. He is a true educator, and it has been my good fortune to work in his laboratory over these past four years.

Table of Contents

Table of Figures

Chapter 1 Introduction

1.1 Motivation and background for systems biology

1.1.1 Signaling networks and protein kinases

Essential to all metazoa is the ability to regulate fundamental cell processes such as death, movement, and replication, thus allowing for the coordination of the vital holistic events that characterize living species, such as development, metabolic processes, and environmental response [1]. Intracellular signaling represents an important mechanism through which individual and groups of cells can regulate these fundamental cell processes. The molecular machinery involved in signaling represents up to 20% of all genes identified by the Human Genome Project [2]. It follows that functional perturbation of these signaling events can result in a host of disease states, with cancer being perhaps the most studied of the resultant illnesses.

The identification of protein serine kinases in the 1960's and protein tyrosine kinases in the early 1980's established these molecules as essential for signal transduction and cell function [3]. In particular, protein tyrosine kinases (PTKs) represent a large fraction of the more than 100 known tumor suppressor genes [2]. Salient PTKs can exist as mobile cytoplasmic or nuclear molecules, membrane bound molecules, or as receptors (RPTKs) that enable communication with extracellular cues such as cytokines and growth factors. In serving as a liaison between extracellular cues and intracellular signaling pathway activation, receptors play a critical role in controlling cell behavior, and as such have become the target of much attention from both the

academic and pharmaceutical groups. The importance of these receptors is clear given the high number implicated in malignancy. Although comprised of a multitude of distinct families, the basic mechanism through which RPTKs transmit signals across cell membranes is quite similar. In general, signal transmission begins with ligand stabilized receptor oligimerization. This then leads to auto and trans-phosphorylation on the cytoplasmic domains of the receptors, which in turn serves to recruit and activate cytoplasmic molecules, many of which are PTKs themselves, that propagate signals throughout the cell. Thus, the final cellular response is a function of the specific activating ligands and the makeup of the intracellular proteins that are phosphorylated or regulated by the RPTK. The basic scheme is outlined below in Figure 1-1 [1].

This thesis explores the mechanisms PTK signaling systems utilize to control normal and misguided cell behavior. We approach the problem from a systems level, rather than a reductionist level, meaning that we hypothesize that cell behavior is regulated through the quantitative manipulation of multiple kinases. Put simply, we have adapted the commonly used DNA→mRNA→protein→cell function paradigm to be DNA→mRNA→protein→protein network→cell function.

That protein kinase networks, and not individual kinases, comprise the basic regulatory unit of biological function is supported by findings from diverse biological fields. These observations have, in many cases, been made with new technologies that allow for network-level measurement. Examples of such technologies include gene arrays, mass spectrometry, short hairpin or short interfering RNA (shRNA or siRNA), and yeast two hybrid assays. For example, a recent shRNA screen performed on human cancer cells identified over 100 candidate gene targets involved in the control of

mitosis and proliferation [4]. Another example is the finding, using a genome-wide siRNA screen against the human kinome, that more than 200 of the approximately 600 kinases screened are involved in cellular endocytosis [5]. Finally, the insufficiency of classical linear signaling pathway description is highlighted in the cell migration of neuronal cells, where two classic but disparately identified pathways, the mitogen-activated protein kinase (MAPK) and phosphatidylinositol 3-kinase (PI3K) pathways, have been shown to both regulate cell migration [6]. High incidence of crosstalk between various pathways, along with spatial and temporal variance, all together create a very confusing view of cell signaling when viewed from a reductionist standpoint. Indeed, in his seminal review of signaling in January of 2000, one of the field's pioneers, Joseph Schlessinger, commented that, "...the modern biochemist and geneticists will have to adopt approaches that have been developed by engineers to describe complicated networks (e.g., systems analysis) in order to obtain a coherent and realistic perspective on cell signaling." [7] Since this review was published, many types of engineering and systems approaches have been applied to biological signaling systems, and as the reader of this thesis will hopefully appreciate in conclusion, these studies have helped to redefine our understanding of how intracellular signaling governs cell function.

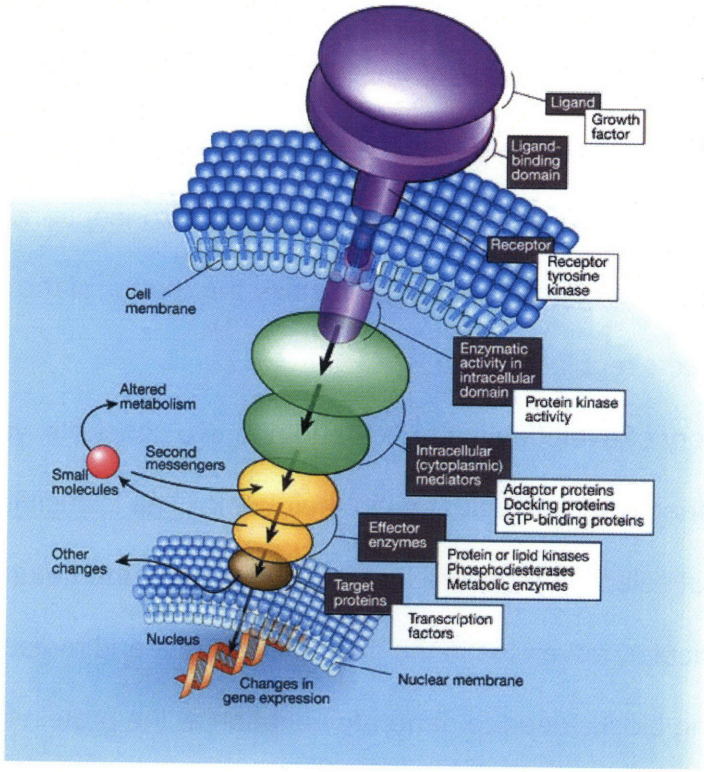


Figure 1-1: A generic signaling pathway. The grey boxes indicate general components of signaling pathways; the white boxes show specific examples [1].

1.1.2 Evolutionary arguments

The application of the systems approach to biology has not only enabled a grater understanding of present-day biology, but has also contributed to our understanding of how cellular life arose. A tenet of systems biology is that a given protein can regulate myriad different phenotypic responses based on variation in the overall signaling network it resides within. This concept has particularly interesting consequences when

viewed in the context of metazoan evolution. Indeed, we learn about why changes in network signaling, versus the creation of multiple separate individual proteins corresponding to various phenotypes, evolved as a strategy for life. In a seminal series of papers, Marc Kirschner and John Gerhardt explore this concept and I highlight some of their findings below in an effort to further convince the reader of the plausibility of systems-level biological regulation.

Darwin's theory of evolution is predicated upon two concepts: variation and selection. As Kirschner and Gerhardt state in their book, The Plausibility of Life, much more is known about selective pressure than is known about the nature of biological variation [8]. In particular, advances in the fields of genetics and molecular biology have identified the building blocks used to generate variation at the protein level (e.g. genes), but the generation of variation at the phenotypic level seems to require the coordinated and combinatorial use of these building blocks, as demonstrated by the low number of genes identified in the human genome (~30,000) and the high degree of conservation between genomes of various species. Rather than viewing the individual gene or gene product as the fundamental block of variance, Kirschner and Gerhardt suggest that we view a collection of core processes (e.g. signaling pathways) as fundamental biological structures that give rise to phenotype. Phenotypic variation, then, can occur through the combinatorial manipulation of these core processes, through the introduction of crosstalk and feedback.

There may exist as few as 17 conserved signaling pathways that have been conserved throughout most metazoa [9]. However, the ability to generate myriad amounts of diversity from these pathways depends on factors such as regulating

temporal dynamics of the pathways, regulating concentrations of different proteins in the pathways, regulating the spatial location of different components of the pathways, and regulating the coordinated control of multiple pathways. In particular, the connection or linkage (Gerhardt and Kirschner refer to this as weak linkage) between proteins in different pathways can give rise to immense amounts of variation. The utilization of core processes to generate variability leads to an increased ability to generate high amounts of non-lethal mutations and phenotypic variation. Part of this ability is related to the inherent robustness associated with regulatory change of conserved pathways, and this is a subject that has been discussed at length elsewhere [8, 10]. In sum, the principle that systems-level signaling controls biological function enables us to understand evolution in the context of the now-detailed characterization of genomic and proteomic elements in metazoa.

1.2 Our approach

To understand how protein networks regulate cellular behavior, we began by abstracting the problem into three different (and obvious) parts, as shown in Figure 1-1. The input space is descriptive of the set of interactions occurring on the extracellular face of the cell. These interactions, such as ligand-receptor interactions, are the first step in information transfer from the extracellular milieu to the internal space of the cell. The signal transfer space is descriptive of information transfer inside of the cell. In particular, we were broadly interested in information disseminated through the transfer of phosphate groups from one molecule to another, a process governed by protein kinases (as already noted). Finally, the response state is descriptive of the cellular

responses such as proliferation, migration, apoptosis, and differentiation. The changing signal state, perturbed via the input state, is somehow responsible for governing cell behavior in a quantitative manner. What does quantitative mean? In terms of cellular response it means that parameters such as the *rate* of proliferation or the *speed* of cellular movement are controlled by the signals, such that these responses can be tuned on a continuous scale, and not just related via an on-off mechanism. The proposition is that the quantitative control of cell response is controlled via quantitative features of cellular proteins, meaning, in particular, that parameters pertaining to kinases of interest, such as signal intensity, on-off rate, duration, and spatial localization, govern eventual cell function. Our reasoning suggests that we must interrogate the signal transfer space in a distributed and quantitative manner. Cell response space must also be measured quantitatively and then mathematical methods to bring the signal and response datasets together must be identified. Furthermore, these measurements should be made under a particular (and hopefully biologically interesting) input condition. Once this input condition is determined, application of experimental and computational procedures, such as those outlined in this thesis, has the potential to provide understanding and suggest therapeutic manipulations of biological systems.

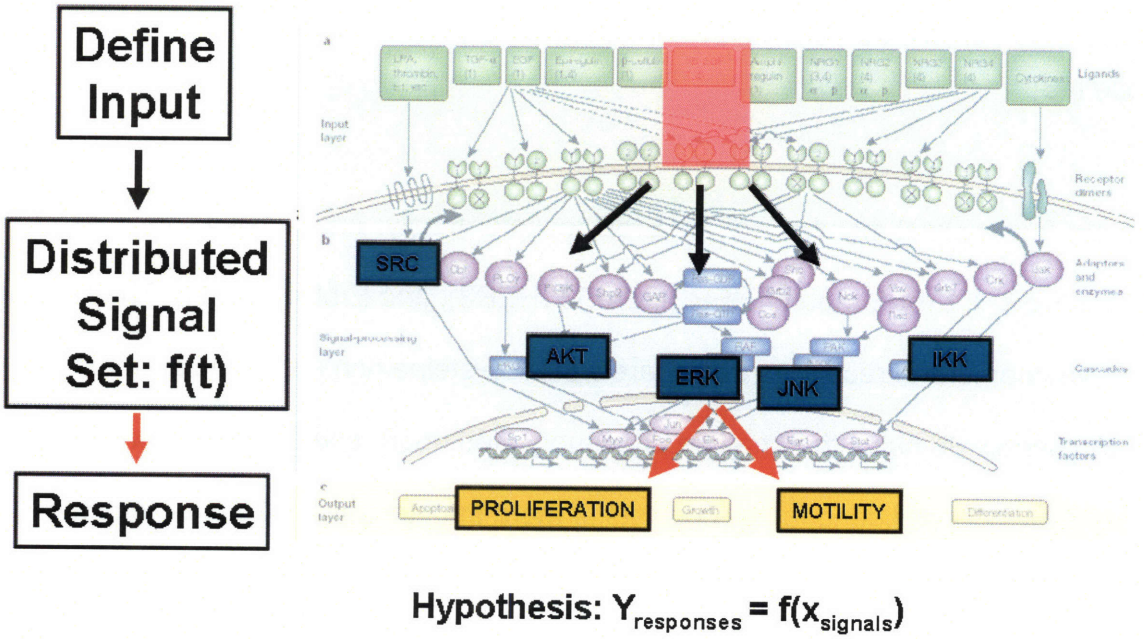


Figure 1-2: Generalized approach to understanding systems control of biological response in terms of intracellular signaling. **Figure adapted from [11].**

1.3 Human epidermal growth factor receptor (HER) background

1.3.1 HER-family receptors and ligands

The EGFR family of receptors and associated ligands has been widely studied. It has served as a model system for the study of PTKs and their biology, as well as one of the most important target systems for the wide variety of cancers associated with PTK deregulation. Indeed, the EGFR or ErbB family of receptors and ligands, has been implicated in 15 types of cancer [11]. In addition to cancer, members of the EGFR family are essential for normal development and function; most ErbB knockouts are non-viable

and ErbB deregulation results in varied pathologies from schizophrenia [12] to cardiac disease [13].

There are four distinct members of the ErbB receptor family: ErbB1 (also known as EGFR or HER1), ErbB2 (neu or HER2), ErbB3 (HER3), and ErbB4 (HER4). They are all type I transmembrane receptors that share similar cysteine rich extracellular domains and cytoplasmic domains that may be phosphorylated upon receptor activation [14]. The ErbB associated ligand family consists of more than ten known ligands, with the salient ones being Epidermal Growth Factor (EGF), transforming growth factor (TGF-α), heparin-binding EGF-like growth factor (HB-EGF), amphiregulin (AR), betacellulin (BTC), epiregulin (EPR), epigen, neuregulin 1 (NRG1 or heregulin), and neuregulins 2-4 (NRG's). Thousands of papers implicate the above ligands in a full physiological spectrum of disease and normal function. Rather than belabor the details of the four different receptors or the many different ligands, we simply state the above to give the reader the sense of the combinatorial complexity that is available for the cell to communicate its various messages through the EGFR system. As shown in figure 1-3, many ligands have the ability to activate unique receptor dimers. One might imagine at its most simplistic level (i.e. not factoring in receptor oligimerization, ligand isoforms, ligand free activation, or simultaneous ligand effects) a combinatorial landscape available to the cell consisting of up to 120 nodes with varying degrees of binding and activation possible at each node, thus allowing for a large number of distinct messages to be transmitted via this specific PTK system (obviously each ligand cannot bind all

receptor pairs and there is downstream redundancy thus making the number of functional nodes available to the system less than indicated but still high).

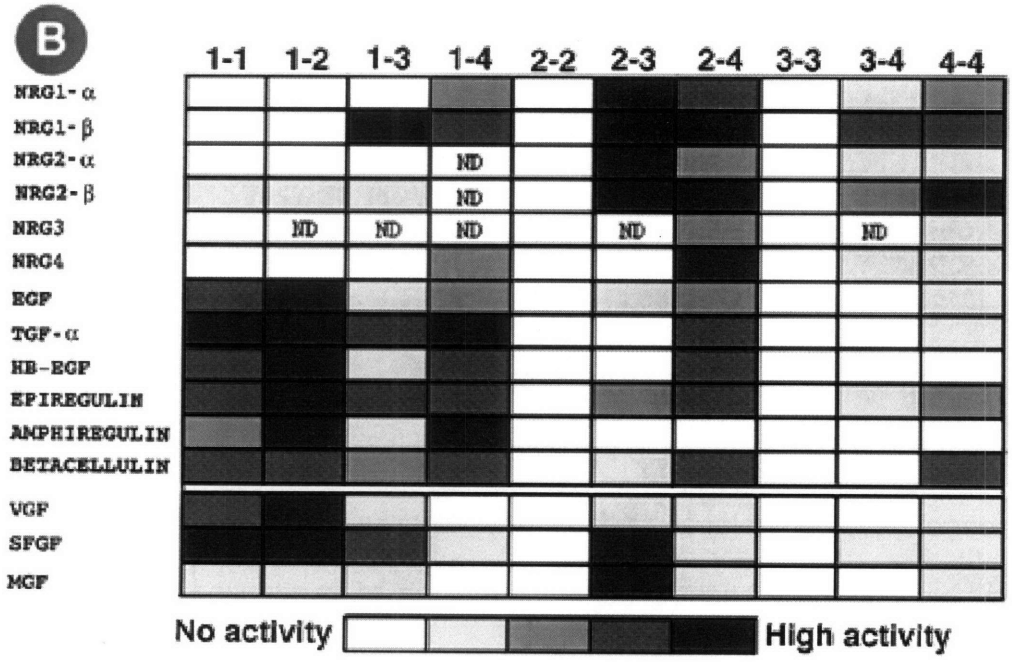


Figure 1-3: **Ligand activation profiles for different ErbB-receptor combinations**. Ten possible dimeric complexes of ErbB receptor proteins are represented (columns). The degree of activation by a particular ligand to a receptor combination is indicated by a shaded scale [15].

Signaling outcome is dictated not just by the combinatorial control available to the cell in terms of ligands and receptor types, but also at the level of kinetic pathway selection. As stated previously, a major goal of this project is to connect signal state information to cellular output. To do this, we must choose a defined set of cellular receptor states such that the problem is tractable and well defined. So, we must increase our magnification of the problem once again. First we moved from the general

framework of PTKs to the more specific system of EGFR. Now, within this EGFR system, we chose to study the specific nature of ErbB2 interactions with EGFR and ErbB3 receptors. In addition, the salient cytokines will be narrowed to EGF and HRG (NRG1).

1.3.3 HER2 biology

ErbB2 overexpression was chosen as the specific perturbing element to look at due to its relevance in breast cancer and other disease states as well as the fact that quantitative studies of ErbB2 and its interactions with EGFR have been performed on cell lines of interest within our lab in the recent past [16, 17]. ErbB2 is critically involved in a number of developmental and normal physiological processes, including cardiac and neural function. ErbB2 has attracted the majority of its attention, however, due to its involvement in breast cancer. Almost 90% of all comedo ductal carcinoma (DCIS) breast cancers feature ErbB2 over-expression which in turn correlates with increased invasiveness, poor patient prognosis, and tumor chemo-resistance [15, 18]. Recent work indicates that ErbB2 is a potent activator of many signaling pathways in over-expressed cell lines. This activation is in part driven through decreased inactivation of signaling complexes resulting from increased recycling and decreased degradation rates. In addition, it has been shown that ErbB2 decreases ligand disassociation rates from receptor complexes [19-21].

Since ErbB2 has no known ligand, heterodimerization is vital to its ability to stimulate downstream signaling events. The most important or well known of ErbB2's

interactions are those with EGFR and ErbB3. In both cases, ErbB2 acts to potentiate signaling complexes. Important also is the fact that ErbB3 lacks tyrosine kinase activity, thus making its interaction with ErbB2 essential to its signaling ability. In the presence of both EGFR and ErbB3, ErbB2 has the option to heterodimerize with either thus stimulating unique signaling paths depending on the distribution of heterodimers between its partners. [22, 23] It has also been shown that ErbB2 will homodimerize, often under receptor over-expressed conditions, in a ligand free manner. Indeed, the general theme of ligand free activation becomes very important when considering the biology of ErbB2 [20]. For instance, although other ErbB receptors (such as EGFR itself) are transforming only in the presence of ligand, ErbB2 is transforming in ligand free environments.

Cell lines stably transfected with specific combinations of ErbB receptors have enabled the study of the individual receptor interactions and their effects on downstream signaling. Recent work has shown that EGFR-ErbB2 interactions in ErbB2 over-expressed cells causes increased activation of the MAPK signaling pathway through ERK [24]. A vast literature linking various heterodimers in the ErbB system to cellular behaviors such as invasiveness, proliferation, and apoptosis exists, but for our purposes here the literature will be used below to motivate the study of specific signaling pathways.

Although HER2 is commonly studied in the context of cancer, it plays an important role under normal physiological conditions. HER2 overexpression varies with tissue type, but is found primarily within epithelial cell layers. The human fetus contains HER2 in the skin, heart, intestine, lungs, and neural system, a finding which implicates

HER2 in embryonic development. Within each tissue, HER2 orchestrates developmental and other such processes through interactions with EGFR family receptors such as ErbB3 and ErbB1, which in turn respond to the specific nature of the activating ligands found in the tissue (with neuregulins and EGF like ligands being examples of such ligands)[25].

Perhaps the most revealing sign that HER2 is vital to embryonic development is the fact that Her2 -/- mice are not viable due to failure in neural and cardiac development.[13] Indeed, even partial rescue of HER2 function using transgenic mice that express the receptor via a cardiac specific promoter (α-MHC promoter) resulted in death at birth, indicating the importance of HER2 on a holistic scale during development.[26] In the neural system, studies indicate that HER2 is involved in the development of the peripheral nervous system (PNS) through the regulation and promotion of Schwann cell migration, acetylcholine receptor clustering, and neuromuscular junction development.[26],[27]. In addition, HER2 seems to play a critical role in the formation of radial glial cells and astrocytes in the cerebral cortex through interplay with the Notch-1 receptor [28]. At the cardiac level, ErbB2 is primarily located in the myocardium, which is composed of muscle cells that regulate the atrium and the ventricle [25]. ErbB2-null mice die due to lack of cardiac traberculae development, pointing to one essential role of HER2 in cardiac development [13]. Other important developmental roles include HER2 regulation of the endothelial-mesenchymal transition, essential for heart valve formation, and transduction of signaling through heparin binding EGF-like growth factors (HB-EGF), which is essential for valve and chamber development [29],[30]. Apart from development, recent work indicates that

HER2 is needed for normal cardiac function, with reports demonstrating that cardiomyopathy results from the directed knockout of HER2 [25]. Furthermore, existing data raises the possibility of a link between known adult HB-EGF deregulation/pathology and HER2.

Studies using Cre/Lox conditional mutant technologies have shown that HER2 is also essential for the development of muscle spindles in skeletal muscle and is required for myoblast cell survival [31],[32]. This data agrees with the previously mentioned findings indicating that HER2 absence deregulates the PNS, since sensory neuron signaling is required for muscle spindle development. HER2 also plays an important role in the normal development of the mammary gland. Specifically, conditional mutation experiments have shown that ErbB2 is required for the formation of lactationally active distended lobuloalveoli [33]. The above results taken together highlight HER2's important role in the development of many tissue types, the normal functioning of developed tissue, and the potential for pathology due to HER2 deregulation.

1.3.4 HER-related signaling pathways

As shown in figure 1-4, there are many signaling pathways activated through the ErbB system. Since the level of complexity is quite high, it is imperative for the experimentalist to choose target signaling molecules and pathways with care so as to maximize the amount of predictive information that can be gained. A wealth of literature is available highlighting ErbB receptor roles in the downstream activation of multiple

signaling pathways. In as much as one can modularize signaling schemes, the ErbB receptor system is implicated in the following pathways: the Shc,Grb2 Ras/MAPK pathway, the MAP3K, MEKK4/7, JNK pathway, the Src kinase family pathway, the JAK/STAT pathway, and the Phospholipid metabolism (PI3-K, PLD, PLC-γ) [34-38]. The following is a brief treatment of some important signaling molecules downstream of the HER-family receptors.

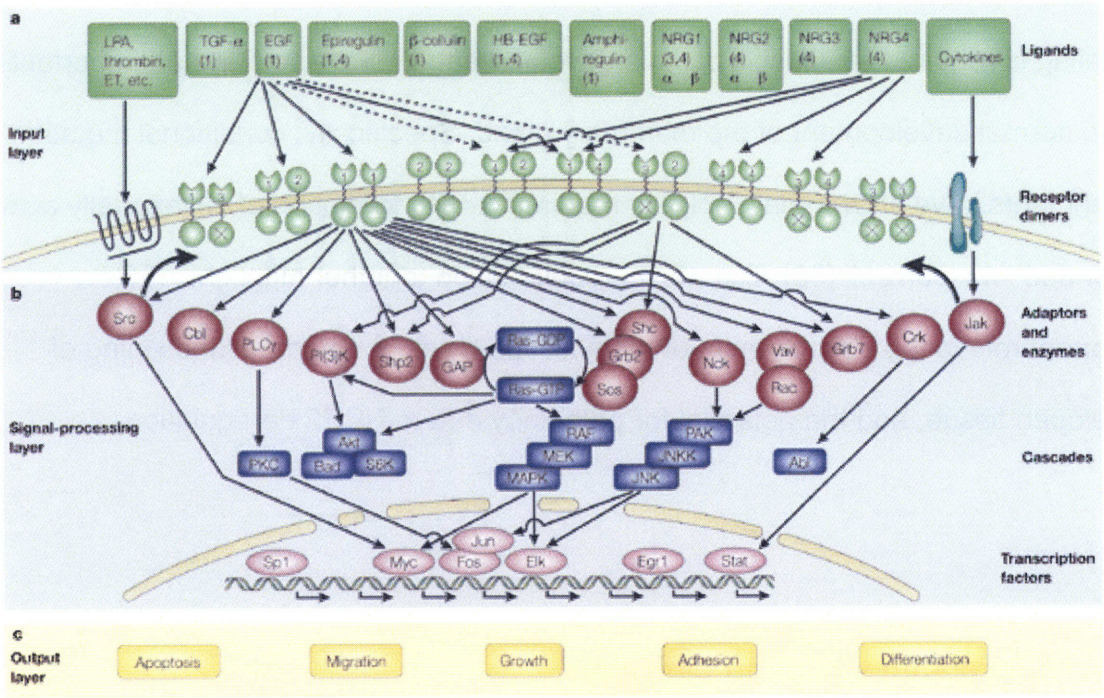


Figure 1-4: A schematic of the ErbB signaling network [11].

Akt(PKB)

The general scheme leading to Akt activation after cytokine binding to a receptor is as follows: the phosphorylated receptor recruits the p85 regulatory subunit of PI3K to the membrane which then recruits the catalytic p110 subunit of PI3K. This activated PI3K complex then phosphorylates PI(4,5)P2 (phosphatidylinositol (4,5)-phosphate) into PI(3,4,5)P3 (phosphatidylinositol (3,4,5)-phosphate). There is also negative regulation at this level through the phosphatases SHIP 1/2 (Src homology 2 containing phosphatases), and PTEN (phosphatase and tensin homologue deleted on chromosome 10), both of which remove phosphate groups from PI(3,4,5)P3 groups. The PIP3 molecules then serve to recruit PDK-1 (phosphotidylinositol-dependant kinase-1) via its PH domain which similarly serves to recruit Akt via its PH domain. PDK-1 phosphorylates Akt at the T308, but full Akt activation seems to require an additional phosphorylation event at S473. The mechanism or kinase that enables this last phosphorylation event has not yet been identified. Akt then goes on to phosphorylate myriad downstream substrates, including Bad, procaspase-9, I-κB (IKK), CREB, the forkhead family of transcription factors, glycogen synthase kinase-3 (GSK-3), Raf, and $p21^{Cip1}$. The Akt pathway has been implicated in many cell responses including proliferation and motility [39]. Specifically within the ErbB family, it has been shown that Akt is often upregulated in ErbB2 over-expressing breast cancers [40, 41] and that this upregulation is linked to a host of cell responses such as increased motility, multidrug resistance, enhanced cell survival, as well as enhanced malignant transformation [40, 42, 43].

JNK

The c-Jun N-terminal kinase (JNK) group of MAP kinases are generally activated by via cellular exposure to stress or cytokine activation. JNK activation is mediated through two MAP kinase kinase (MKK) molecules, MKK4 and MKK7. These are in turn regulated by MAP kinase kinase kinases (MEKKs) such as MEKK1, MUK/DLK, and others. JNK activation is a highly complex process that can be mediated by disparate molecules such as PI3K and Rho GTPases. JNK protein kinases have been shown to regulate transcription through the phosphorylation of transcriptional activation domains ATFa, c-Jun, JunD, Elk-1, and Sap-1. The resultant cell behavior due to JNK activity has been far less well characterized. Indeed JNK's role in varied responses from apoptosis to inflammation has been widely debated in the literature and may be extremely cell type dependant. Furthermore, its role in tumor progression is very poorly understood [44-46]. Much of the confusion may have to do with the fact that many signaling pathways converge upon the MAPK pathway at the JNK level, leading to a high degree of complexity and redundancy at the upstream level. JNK is an ideal molecule for our study not only because it seems to be critically involved in many cell processes, but also because its complex responses may prove to be ideal for model analysis. Cell motility has recently been shown to be regulated by JNK in some systems. JNK1 activity seems to be essential for rapid cell migration in rat bladder epithelial cells (NBT-II) [47]. Previous work has heavily implicated MEK kinase 1 (MEKK1), an upstream regulator of the JNK pathway, in cell motility [48]. In sharp contrast to the PI3K and ERK pathways, very little has been published about JNK's involvement in ErbB2 over-expressing systems, although there is a small body of literature that suggests it may be important for malignant transformation [49].

ERK

The ERK MAPK pathway is one of the most studied cytokine activated pathways. The general scheme of activation through ErbB family receptors is as follows: active phosphorylated receptors bind Grb2 directly or indirectly through Shc, thus recruiting Sos, a guanine nucleotide exchange factor for Ras. This then promotes Ras activation which interacts with the serine-threonine kinase Raf-1. Raf-1, through a number of intermediate molecules such as MEK, then activates ERK which goes on to phosphorylate a number of cytoplasmic substrates as well as transcription factors in the nucleus. There is crosstalk between this path and the PI3K pathway at the level of Raf-1, as well as considerable crosstalk between this and the JNK pathway at the level of the AP-1 transcriptional unit. Thus, once again, ERK becomes an ideal molecule to study and model due to its complex integration of signaling activation. ERK has been heavily implicated in ErbB family signaling [25, 34]. It has been shown to regulate cell motility and proliferation through a variety of different mechanisms [34, 50-53].

c-Src

Src kinases are essentially membrane bound molecular switches that serve to coordinate receptor changes and intracellular pathways. Src has been implicated in a wide variety of cell behaviors such as cell cycle control, proliferation, motility, and differentiation [54, 55]. The Src kinases are activated by a variety of different surface receptors that serve often to bind the SH2 domain of Src, thus releasing it from its naturally auto-inhibited state. ErbB family interactions with Src have been well documented in the case of EGFR. C-Src binds to EGFR and contributes to a number of

downstream signaling events. Substrates of c-Src include many proteins that are critical in cytoskeletal reorganization events and motility, such as FAK, p130Cas, cortactin, EAST, and Eps-8 [56-58]. In addition, Src kinases have been shown to activate the PI3K pathway [59]. Thus, here again we have a molecule that is linked to multiple pathways and cellular behaviors, thus making it an ideal target for study via the modeling approach.

1.4 Human mammary epithelial cell (HMEC) system

1.4.1 Cell line background and considerations

Cell line models represent a critical parameter of choice. One of the best characterized cell lines in relation to ErbB activity is the 184A1 cell human mammary epithelial cell line. The 184A1 cell line was initially characterized by Band and Sanger [60] and subsequently analyzed by numerous groups [17, 20]. The cell line has subsequently been engineered to parse out the interactions between ErbB2 and EGFR. Toward that goal, four clones of the 184A1 cell line have been produced using retroviral introduction of the ErbB2 gene. These four clones represent a model system for epithelial cells having varying levels of ErbB2 on the surface. Table 1 highlights the receptor number information for this cell line. Work with these models clones has yielded the most comprehensive quantitative set of data available with respect to ErbB2-EGFR interactions. Quantitative models for both receptor dimerization and trafficking phenomena have also been generated using data from this cell line [16, 17].

It has, however, been shown that the 184A1 cell line is generally non-tumorigenic in nude mice, as noted further in the next section. A thorough analysis of current literature and cell lines models of interest in pharmaceutical companies indicates that the below cell lines, while not used in our study, also may serve as appropriate models for future studies.

The H16N2 human mammary epithelial cell line has been manipulated using retroviral techniques to yield over-expressing ErbB2 cells. Over-expression in this cell line has been implicated in increased invasiveness and activation of the PI3K pathway [42]. Another cell line of choice is the MDA-MB family of cell lines. In particular, the MDA-MB-435 and the MDA-MB-435/Her-2 (stable over-expression of HER2 from transfection) offer an excellent model cell line in which to test effects of HER2 over-expression. Studies on this cell line to date have shown that HER-2 over-expression confers enhanced cell survival and increased Akt activity. The same qualities have been shown to be true for the MCF7 and the MCF7/HER-2 over-expressing cell line [40]. A recent paper by Knuefermann et. al. shows that PI3K mediated activation of Akt is found predominantly in cells over-expressing both ErbB2 and ErbB3, such as the MCF7, the MDA453, and the MDA361. This is in direct contrast to those that have high levels of only one receptor, such as MDA231 (EGFR), BT474 or SKBR3 (ErbB2), and MDA435 (ErbB3) [43]. This then allows for validation of observations made in the MDA435 case, with increased ErbB2 levels, against some natural cell lines, such as the MDA453. In sum, use of the MDA cell line family along with retroviral and knockdown techniques to obtain a working model of ErbB2/ErbB3 interaction should permit observation of behaviors relevant to breast cancer.

Cell Line	EGFR Number	ErbB2 Number
184A1 HMEC (Parental)	~$2x10^5$	~$2x10^4$
A1-1 clone 29L	~$2x10^5$	~$1x10^5$
A1-1 clone 12	~$2x10^5$	~$2x10^5$
A1-1 clone 24H	~$2x10^5$	~$6x10^5$

Table 1-1: EGFR and HER2 receptor levels for the engineered 184A1 HMEC cell line

1.4.2 Characteristics of HMEC cells

The human mammary epithelial cell line 184A1 was used throughout the course of this thesis to study the effects of HER2 overexpression on cell signaling and behavior. A knowledge of cell source and the method of immortalization is necessary when considering how accurate the cell line is for predictions *in vivo* and how it compares to results in other breast epithelial cell lines. In general, human mammary epithelial cells are procured through reduction mammoplasties or mastectomies. In addition, these cells can be obtained from lactational fluids and needle aspirations [61]. The HMEC 184A1 cells were derived from a reduction mammoplasty and then immortalized through exposure to the chemical carcinogen benzo(a)pyrene [60-62]. These cells have homozygous mutations of the cyclin dependent kinase inhibitor (CKI) p16 gene but stable levels of p53 [61]. Importantly, HMEC 184A1 cells show low genomic instability,

allowing for longer-term passaging as compared to breast cancer epithelial cells. 184A1 express polymorphic epithelial mucins (PEM), keratins 5/14 and 8/18, but not express estrogen receptor (ER) [61, 62]. The cells are non-tumorigenic in nude mice and are growth factor and anchorage dependent [62]. 24H cells, or HMEC 184A1 cells stably transfected with HER2, are also non-tumorigenic and anchorage and growth factor dependent, although it is not known how expression of PEM or keratins change in these cells. Finally, it is interesting to note that HMEC 184A1 cell lines undergo a slow nonuniform growth phase for the first 20-30 passages, and the HMEC experiments described throughout the course of the work were performed on 184A1 that had been passaged past this point [62]. There is no known mechanistic reason for this early nonuniform growth phase. Thus, the HMEC 184A1 system provides a well-characterized system in which to study the effects of HER2 overexpression in normal breast epithelia.

1.5 Computational Models

1.5.1 A brief introduction to computational modeling in biology

The coordinated regulation of biological signaling has many of the same properties we observe in other engineering systems, e.g. feedback, amplification, and crosstalk. As the number of molecules and interactions that govern one cellular function becomes large, it becomes difficult to intuit the underlying rules or principles that govern

the system. Computational models help to codify potential intuition from highly complex datasets, allowing for hypothesis generation and the identification of multi-dimensional signaling control elements for cell behavior. Introduction to the many types of computational models is beyond the scope of this thesis. However, I offer a broad overview of three important classes of computational models in Chapter 8, with special attention to their usefulness for prediction in the pharmaceutical field. In addition, I would refer the interested reader to the following references for more general reviews of cell signaling models [63-68].

In particular, this thesis is concerned with models that describe and predict cellular output from network signals. Excellent reviews of this class of models are available in [66, 69] as well as in Chapter 8. Briefly, these models use diverse computational techniques, for example linear mapping and decision tree analysis, to understand how measured intracellular signaling events (typically phosphorylation changes) govern changes in cell behavior (such as cell migration or apoptosis).

Although many excellent reviews of various computational approaches exist as mentioned above, the first step in approaching the problem of how signals govern behavior is to understand the dimensionality or scope of the problem. As engineers, we typically address this issue by taking stock of the orders of magnitude inherent in the problem. Here, I briefly explore these orders of magnitude to highlight two important points: 1) the complexity of the problem, which emphasizes the need form computational aid in understanding and 2) the inability of any one model to capture all processes, thus emphasizing the need for careful model assumptions and

simplifications. Many of the below estimates are taken from an excellent review by Papin et al. [68].

There are ~30,000 genes in the human genome encoding ~500 protein kinases, ~150 protein phosphatases, and ~1,500 receptors (Table 1). Furthermore, each gene has a conservative average of ~3 splice variants and further ~3 post-translational modifications (e.g. phosphorylation or methylation). Thus, the number of different states achievable as defined by the collection of final protein products is in excess of 50,000. As discussed in Chapter 1.1.2, further linkage or protein-protein interaction also occur, with an average of ~5 interactions per protein. Thus the number of possible states achievable in a signaling network at a given period of time is immense. Clearly, if we suggest that understanding network properties is essential to understanding the governance of cell behavior, the problem is much more complicated than human intuition can account for, but also more complicated than current computational models can account for. Thus, the appropriate abstraction of complexity becomes necessary, and, as noted elsewhere, depends on the questions being posed in the study and the nature of the data available [65, 66].

Table 1 | **The scope of the human cellular signalling network**

Network component	**Number**
Cells	10^{14}
Cell types	200
Genes	25,000
Percentage of genes with splice variants	40–60
Average number of exons per alternatively spliced gene	8
Maximum number of exons per alternatively spliced gene (taken from *TTN* that encodes titin)	234
Average number of splice variants per gene across the genome	2.5
Percentage of alternatively spliced genes with signalling function	75
Average number of post-translational modifications per protein (current estimates)	2.5
Genes for transcription factors	1,850
Genes for protein kinases	518
Genes for protein phosphatases	150
Genes for receptors	1,543
Genes for GPCRs (for endogenous ligands)	367

*See the Human Proteomics Initiative in the online links box.

Table 1-2: Orders of magnitude for parameters in cell signaling networks. Taken from [68].

Our above discussion omitted added complexity arising from the dynamics of signaling networks. Signaling dynamics play a key role in determining resultant cell function as mentioned in Chapter 8. Although signaling and protein modification can occur on the order of microseconds, transcription, later-phase signaling, and cell function occur across diverse timescales (Table 2), indicating that careful thought as to how to integrate models across timescales must be paid prior to analysis [70]. Thus, taking stock of the parameter space involved in understanding biological signaling and behavior reveals a highly complicated system that requires the carefully focused use of computational models.

Cellular signalling process	Time (in seconds)
Activities	
Kinase/phosphatase reactions	10^{-3}
Protein conformational changes	10^{-3}
Cell-scale protein diffusion (passive)	10^{0}–10^{1}
Cell-scale protein diffusion (active)	$<10^{0}$
Responses	
Cell migration	10^{0}–10^{2}
Receptor internalization	10^{2}
Transcriptional control	10^{2}
Cellular growth	10^{4}

Table 1-3: Relevant orders of magnitude for timescales in network-function models. Taken from [68].

1.5.1 Partial Least Squares Regression (PLSR): An introduction

In Chapters 4 and 5, I introduce the specifics of the modeling approach used in this thesis, namely, partial least squares regression (PLSR). PLSR is a method that enables the linear mapping of one dataset onto another (i.e. signals onto cell function), and has at its heart two important concepts: linear regression and single value decomposition (SVD). In this section, I discuss the tenets of the linear approach we undertook, and then introduce PLSR by way of SVD and linear regression. The details of our methodology are further explained in later chapters as noted throughout the following discussion.

When considering how a set of signals relates to cellular response or behavior, it is difficult to intuit the form that the function would take. For instance, if one is looking at cell migration, one might posit that migration = f(Erk, Akt). However, there is not enough underlying data about the chain of chemical reactions that lead to Erk's or Akt's effects

on cell migration for us to explicitly enumerate such as function. Hence, approximating functions must be used. As explained in Chapter 8, a high level of abstraction is generally needed when considering cell signal-cell response type models. Linear mapping is a highly abstracted computational tool ideal for signal-response analysis since we can then take advantage of the tools linear algebra provides as well as the rigorous statistical theory developed around linear regression. The general problem, as portrayed below in Figure 1-5, is defining a set of parameters that project a matrix of signals onto a matrix of cell behavior measured under the same cellular conditions. We define the set of parameters that transform the signals to behavior (denoted **b** in later equations) by performing a linear regression. The hypothesis behind the work presented in this thesis is that the linear projection defined by **b** will reveal underlying principles and rules governing the coordinated control of cell behavior by many kinases, as well as having a sufficiently large range such that application to new signaling sets generated after model completion (e.g. those corresponding to addition of new inhibitor, ligand, or receptor expression) will accurately compute novel cell behavior *a priori*.

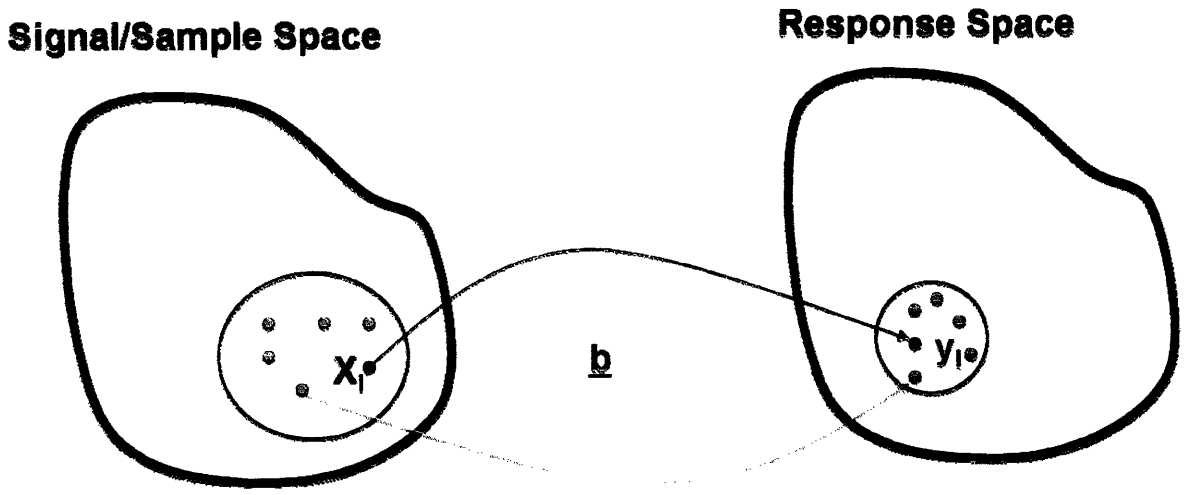


Figure 1-5: **Mapping signals to cellular response using linear projection.** The vector **b** takes a matrix of signals (**X**) and projects it to a set of responses (**Y**). The row and column spaces are shown as subspaces within the signal and response spaces, respectively.

SVD, a critical component of the PLSR algorithm, is a mathematical technique with a rich history in the field of linear algebra as well as usefulness in the field of biological study [71, 72]. Indeed, the noted linear algebra professor Gilbert Strang called SVD "a high point of linear algebra" in his book, Introduction to Linear Algebra [73]. SVD is a powerful biological approach due to the high incidence of rank-deficient matrices in biological modeling. The matrix of signaling values in the work presented herein, which I shall call **X**, is an *MxN* matrix corresponding to *M* cellular conditions measured and *N* protein measurements for each condition. A row in **X** might refer to a particular cell type stimulated with EGF, whereas a column refers to the measurement of Akt phosphorylation or activity at 5 minutes under various cell conditions. Typically, when utilizing technologies that measure kinase properties from whole-cell lysates, we have many more columns than we do rows ($N>M$), thus rendering our matrix rank deficient. The number of columns can be quite large (see Chapters 4 and 5), often in excess of 1000, whereas the number of cell conditions is typically under 20. To reduce the dimensionality of the signaling matrix and to identify an orthogonal basis set for both the row and column spaces of **X**, we require SVD, which decomposes the matrix as follows:

$$X = U \sum V^T \tag{1}$$

where the first *r* columns of **V** are an orthogonal basis set for the row space and the first *r* columns of **U** are an orthogonal basis set for the column space, with *r* being the rank of matrix **X**. Columns after *r* in **V** and **U** correspond to orthogonal bases for the null spaces of **X** and $\mathbf{X}^T$, respectively. With this decomposition, we can write expressions for the covariance of both X and X^T as follows:

$$XX^T = U\sum\sum^T U^T \tag{2}$$

$$X^T X = V\sum^T \sum V^T \tag{3}$$

Since these equations are in Jordan normal form, the columns of **V** are the eigenvectors for $\mathbf{X}^T\mathbf{X}$ and the columns of **U** are the eigenvectors for $\mathbf{X}\mathbf{X}^T$. Thus, to solve for **U** and **V**, we need to calculate $\mathbf{X}^T\mathbf{X}$ and $\mathbf{X}\mathbf{X}^T$ and then solve for their eigenvectors.

Computationally, we use an efficient technique to calculate both **U** and **V** called non-linear iterative partial least squares (NIPALS). Before discussing this algorithm, I define the terms used in the algorithm. The need for further definition arises from the fact that SVD is also known as principal components analysis (PCA), and although these two names refer to the same technique, different terms (and typically different derivation and explanation methods) are used in the PCA literature. Briefly, papers in the PCA field typically decompose a matrix into a sum of outer products as follows [74]:

$$X = t_1 p_1^T + t_2 p_2^T + ...t_r p_r^T \tag{4}$$

The vector *t* is the scores vector and represents the columns in the **U** matrix defined during our SVD discussion. The vector *p* is the loadings vector and corresponds to **V** from our SVD discussion. These vectors are often represented graphically as shown below in Figure 2.

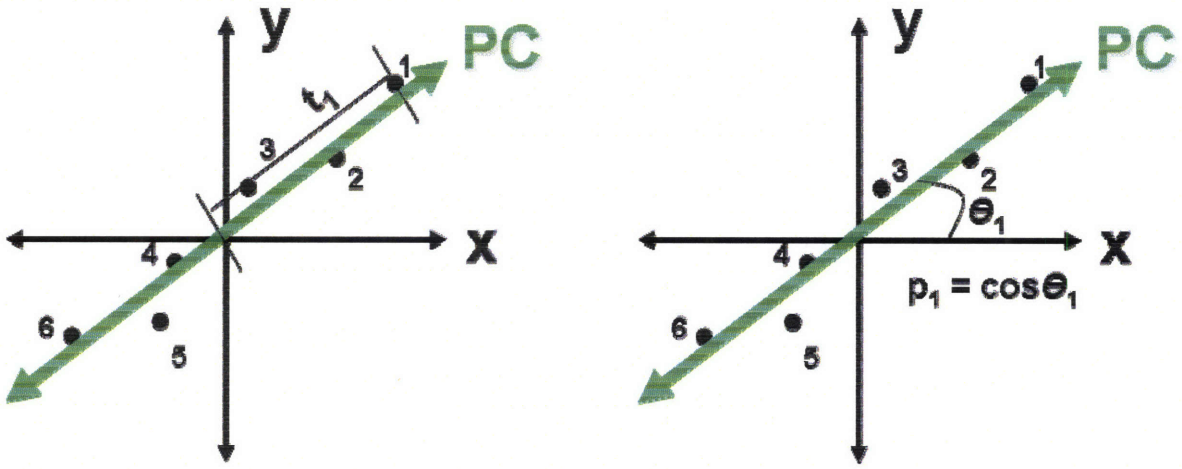


Figure 1-6: **A graphical representation of PCA in two dimensions.** Here the PC is the line of best fit and the scores are the projection of individual data along the direction of the PC (right figure) whereas the loadings are the cosine angles of the direction vector with respect to the original axes.

The NIPALS algorithm, then, as stated in [74], for the calculation of PC_h is:

1) take vector x_j from X and define it as t_h

2) calculate p_h^T by solving $p_h^T = t_h^T X / t_h^T t_h$

3) normalize p_h^T to unit length

4) calculate new t_h by solving $t_h = X p_h / p_h^T p_h$

5) compare t_h used in step 2 to the one generated in step 4. If they are the same within some predetermined level of accuracy, then stop. If not, then go back to step 2 for another iteration.

The NIPALS algorithm is the same eigenvalue problem discussed when introducing SVD, as can be seen by combining the equations in steps 2 and 4 to get the following eigenvalue problems:

$$0 = t_h\left(C_1 I - XX^T\right) \tag{5}$$

$$0 = p_h^T\left(C_1 I - X^T X\right) \tag{6}$$

Where C_1 is $\left(p_h^T p_h\right)\left(t_h^T t_h\right)$. Thus, the NIPALS algorithm can be used to generate the single value decomposition or the principal components of a matrix.

We now return to our original task of defining the vector **b** for the projection of our signaling matrix to our cellular response space. The reason for introducing SVD first is because we need it to solve the linear regression problem stated in equation 7 when dealing with an **X** that is rank-deficient.

$$b = \left(X^T X\right)^{-1} X^T y \tag{7}$$

We can use SVD to rewrite the problem using the psuedoinverse, defined as $\mathbf{X^{-1}}$, by:

$$b = X^{-1} y = V \Sigma^{-1} U^T y \tag{8}$$

We can also decompose **Y** to rewrite the above equation in terms of the two SVD's. The decomposition of both the **X** and **Y** matrices to perform linear regression is called principle component regression (PCR). It is important not only because it allows us to

solve the linear problem with rank-deficient matrices, but also because as the number of variables rise in a linear model, the uncertainty in the estimated parameters rises in parallel and becomes the dominating factor [75]. Thus, by performing SVD on both the **X** and **Y** matrices, we reduce the dimensionality associated with each matrix, thereby increasing the predictive power of the resulting model. The reduction in number of variables is due to the fact that regression parameters are determined in PCR between the scores vectors of the **X** and **Y** matrices. If we represent both **X** and **Y** by their associated scores vectors, *t* and *w*, respectively, we obtain the following regression equation:

$$t\gamma = w\,; \quad \gamma = \frac{p^T bq}{q^T q} \tag{9}$$

Where *p* and *q* are the loadings vectors associated with **X** and **Y**, respectively. It is natural to represent **X** and **Y** by their scores vectors since these vectors represent an appropriately 'blended' column space, or put more precisely, the basis set for the covariance of each matrix. Thus, by decomposing our matrices using SVD and then performing regression to obtain **b**, we have solved the linear projection problem we posed at the outset of this chapter. One final problem remains, however. And that is the fact that by decomposing the **X** and **Y** matrices separately, we did not define principal components that necessarily best captured the covariance *between* the two matrices, which is the covariance we wish to model. To do this, we define principal components such that we account for the covariance between X and Y using a procedure called PLSR. This procedure is implemented using a variation of the NIPALS algorithm, and we describe it in greater detail in the Methods sections of Chapter 5. Here, we present the algorithm for the PLSR NIPALS approach, as documented in [75], and then define

the PLSR loadings and scores vectors in terms of an eigenvalue problem, highlighting the additional accounting for covariance between the **X** and **Y** matrices. The PLSR NIPALS algorithm is:

1) define *w* as the first column of **Y**

2) $a = X^T w/(w^T w)$

3) Scale *a* to unit length.

4) $t = Xa$

5) $c = Y^T t/(t^T t)$

6) Scale *c* to be unit length.

7) $w = Yc/(c^T c)$

8) check *w* in Step 7 against that of Step 2. If they are equal, the continue to the next step, if not, return to Step 2.

9) $p = X^T t/(t^T t)$

10) $q = Y^T w/(w^T w)$

11) perform regression: $\gamma = w^T t/(t^T t)$

Note that here we have two vectors, *a* and *c*, that define the loadings associated with variance-covariance matrices as defined below, whereas *p* and *q* continue to relate directly to **X** and **Y** only. The variables *t, w, a, and c* correspond to the following eigenvectors [75]:

$$w \rightarrow YY^T XX^T$$

$$c \rightarrow Y^T XX^T Y$$

$$t \rightarrow XX^T YY^T$$

$$a \rightarrow X^T Y Y^T X$$

Thus, using PLSR, we solve a linear regression problem and account for covariance between our signals and cellular behavior to define a projection that linearly describes and predicts cellular decision making processes.

1.6 Thesis Overview

The work described in this thesis is a small part of the movement toward a more predictive form of biological science. The product of the work is not only the development and application of models focused on understanding how protein networks regulate cell function in important cancer systems, but also a description of how a combination of experimental and computational approaches can lead an investigator toward more focused biological problems. As such, although a majority of the work described below deals with the data capture and model development toward the understanding of signal-response behavior, Chapters 2 and 7 contain more detailed studies into the biophysical and biochemical mechanisms behind cell migration and Akt activation. Both topics arose naturally from the systematic evaluation of the signaling networks in cancer cells. Chapters 2 and 3 together describe the measurement techniques used to gather quantitative cell signal and cell behavior data. Chapters 4 and 5 describe the analysis of gathered quantitative data in HER2 overexpressing cells using linear modeling. Chapter 6 explores the network signaling properties of small molecule inhibitors, and combines both experimental and modeling approaches to predict inhibitor efficacy against HER2-mediated migration. Chapter 7 contains the

above-described analysis of Akt signaling and Chapter 8 concludes the thesis by presenting potential applications of computational models in the pharmaceutical industry.

References

1. Downward, J., *The ins and outs of signalling.* Nature, 2001. **411**(6839): p. 759-62.
2. Blume-Jensen, P. and T. Hunter, *Oncogenic kinase signalling.* Nature, 2001. **411**(6835): p. 355-65.
3. Hunter, T., *Signaling--2000 and beyond.* Cell, 2000. **100**(1): p. 113-27.
4. Moffat, J., et al., *A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen.* Cell, 2006. **124**(6): p. 1283-98.
5. Pelkmans, L., et al., *Genome-wide analysis of human kinases in clathrin- and caveolae/raft-mediated endocytosis.* Nature, 2005. **436**(7047): p. 78-86.
6. Segarra, J., et al., *Combined signaling through ERK, PI3K/AKT, and RAC1/p38 is required for met-triggered cortical neuron migration.* J Biol Chem, 2006. **281**(8): p. 4771-8.
7. Schlessinger, J., *Cell signaling by receptor tyrosine kinases.* Cell, 2000. **103**(2): p. 211-25.
8. Kirschner, M. and J. Gerhart, *The Plausibility of Life: Resolving Darwin's Dilemma*. 2005, New Haven, CT: Yale University Press.
9. Gerhart, J., *1998 Warkany lecture: signaling pathways in development.* Teratology, 1999. **60**(4): p. 226-39.
10. Kitano, H., *Biological robustness.* Nat Rev Genet, 2004. **5**(11): p. 826-37.
11. Yarden, Y. and M.X. Sliwkowski, *Untangling the ErbB signalling network.* Nat Rev Mol Cell Biol, 2001. **2**(2): p. 127-37.

12. Falls, D.L., *Neuregulins: functions, forms, and signaling strategies.* Exp Cell Res, 2003. **284**(1): p. 14-30.

13. Lee, K.F., et al., *Requirement for neuregulin receptor erbB2 in neural and cardiac development.* Nature, 1995. **378**(6555): p. 394-8.

14. van der Geer, P., T. Hunter, and R.A. Lindberg, *Receptor protein-tyrosine kinases and their signal transduction pathways.* Annu Rev Cell Biol, 1994. **10**: p. 251-337.

15. Harari, D. and Y. Yarden, *Molecular mechanisms underlying ErbB2/HER2 action in breast cancer.* Oncogene, 2000. **19**(53): p. 6102-14.

16. Hendriks, B.S., et al., *Quantitative analysis of HER2-mediated effects on HER2 and epidermal growth factor receptor endocytosis: distribution of homo- and heterodimers depends on relative HER2 levels.* J Biol Chem, 2003. **278**(26): p. 23343-51.

17. Hendriks, B.S., et al., *Coregulation of epidermal growth factor receptor/human epidermal growth factor receptor 2 (HER2) levels and locations: quantitative analysis of HER2 overexpression effects.* Cancer Res, 2003. **63**(5): p. 1130-7.

18. Holbro, T., G. Civenni, and N.E. Hynes, *The ErbB receptors and their role in cancer progression.* Exp Cell Res, 2003. **284**(1): p. 99-110.

19. Lenferink, A.E., et al., *Differential endocytic routing of homo- and hetero-dimeric ErbB tyrosine kinases confers signaling superiority to receptor heterodimers.* Embo J, 1998. **17**(12): p. 3385-97.

20. Worthylake, R., L.K. Opresko, and H.S. Wiley, *ErbB-2 amplification inhibits down-regulation and induces constitutive activation of both ErbB-2 and epidermal growth factor receptors.* J Biol Chem, 1999. **274**(13): p. 8865-74.

21. Wada, T., X.L. Qian, and M.I. Greene, *Intermolecular association of the p185neu protein and EGF receptor modulates EGF receptor function.* Cell, 1990. **61**(7): p. 1339-47.

22. Olayioye, M.A., et al., *ErbB-1 and ErbB-2 acquire distinct signaling properties dependent upon their dimerization partner.* Mol Cell Biol, 1998. **18**(9): p. 5042-51.

23. Graus-Porta, D., R.R. Beerli, and N.E. Hynes, *Single-chain antibody-mediated intracellular retention of ErbB-2 impairs Neu differentiation factor and epidermal growth factor signaling.* Mol Cell Biol, 1995. **15**(3): p. 1182-91.

24. Hendriks, B.S., In Preparation, 2003.

25. Citri, A., K.B. Skaria, and Y. Yarden, *The deaf and the dumb: the biology of ErbB-2 and ErbB-3.* Exp Cell Res, 2003. **284**(1): p. 54-65.

26. Morris, J.K., et al., *Rescue of the cardiac defect in ErbB2 mutant mice reveals essential roles of ErbB2 in peripheral nervous system development.* Neuron, 1999. **23**(2): p. 273-83.

27. Lin, W., et al., *Aberrant development of motor axons and neuromuscular synapses in erbB2-deficient mice.* Proc Natl Acad Sci U S A, 2000. **97**(3): p. 1299-304.

28. Schmid, R.S., et al., *Neuregulin 1-erbB2 signaling is required for the establishment of radial glia and their transformation into astrocytes in cerebral cortex.* Proc Natl Acad Sci U S A, 2003. **100**(7): p. 4251-6.

29. Camenisch, T.D., et al., *Heart-valve mesenchyme formation is dependent on hyaluronan-augmented activation of ErbB2-ErbB3 receptors.* Nat Med, 2002. **8**(8): p. 850-5.

30. Iwamoto, R., et al., *Heparin-binding EGF-like growth factor and ErbB signaling is essential for heart function.* Proc Natl Acad Sci U S A, 2003. **100**(6): p. 3221-6.

31. Andrechek, E.R., et al., *ErbB2 is required for muscle spindle and myoblast cell survival.* Mol Cell Biol, 2002. **22**(13): p. 4714-22.

32. Leu, M., et al., *Erbb2 regulates neuromuscular synapse formation and is essential for muscle spindle development.* Development, 2003. **130**(11): p. 2291-301.

33. Jones, F.E. and D.F. Stern, *Expression of dominant-negative ErbB2 in the mammary gland of transgenic mice reveals a role in lobuloalveolar development and lactation.* Oncogene, 1999. **18**(23): p. 3481-90.

34. Jorissen, R.N., et al., *Epidermal growth factor receptor: mechanisms of activation and signalling.* Exp Cell Res, 2003. **284**(1): p. 31-53.

35. Belsches, A.P., M.D. Haskell, and S.J. Parsons, *Role of c-Src tyrosine kinase in EGF-induced mitogenesis.* Front Biosci, 1997. **2**: p. d501-18.

36. Darnell, J.E., Jr., *STATs and gene regulation.* Science, 1997. **277**(5332): p. 1630-5.

37. Kamat, A. and G. Carpenter, *Phospholipase C-gamma1: regulation of enzyme function and role in growth factor-dependent signal transduction.* Cytokine Growth Factor Rev, 1997. **8**(2): p. 109-17.

38. Carpenter, C.L., et al., *Phosphoinositide 3-kinase is activated by phosphopeptides that bind to the SH2 domains of the 85-kDa subunit.* J Biol Chem, 1993. **268**(13): p. 9478-83.

39. Chang, F., et al., *Involvement of PI3K/Akt pathway in cell cycle progression, apoptosis, and neoplastic transformation: a target for cancer chemotherapy.* Leukemia, 2003. **17**(3): p. 590-603.

40. Bacus, S.S., et al., *AKT2 is frequently upregulated in HER-2/neu-positive breast cancers and may contribute to tumor aggressiveness by enhancing cell survival.* Oncogene, 2002. **21**(22): p. 3532-40.

41. Stal, O., et al., *Akt kinases in breast cancer and the results of adjuvant therapy.* Breast Cancer Res, 2003. **5**(2): p. R37-44.

42. Ignatoski, K.M., et al., *ERBB-2 overexpression confers PI 3' kinase-dependent invasion capacity on human mammary epithelial cells.* Br J Cancer, 2000. **82**(3): p. 666-74.

43. Knuefermann, C., et al., *HER2/PI-3K/Akt activation leads to a multidrug resistance in human breast adenocarcinoma cells.* Oncogene, 2003. **22**(21): p. 3205-12.

44. Davis, R.J., *Signal transduction by the JNK group of MAP kinases.* Cell, 2000. **103**(2): p. 239-52.

45. Ip, Y.T. and R.J. Davis, *Signal transduction by the c-Jun N-terminal kinase (JNK)--from inflammation to development.* Curr Opin Cell Biol, 1998. **10**(2): p. 205-19.

46. Johnson, G.L. and R. Lapadat, *Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases.* Science, 2002. **298**(5600): p. 1911-2.

47. Huang, C., et al., *JNK phosphorylates paxillin and regulates cell migration.* Nature, 2003. **424**(6945): p. 219-23.

48. Yujiri, T., et al., *MEK kinase 1 gene disruption alters cell migration and c-Jun NH2-terminal kinase regulation but does not cause a measurable defect in NF-kappa B activation.* Proc Natl Acad Sci U S A, 2000. **97**(13): p. 7272-7.

49. Amundadottir, L.T. and P. Leder, *Signal transduction pathways activated and required for mammary carcinogenesis in response to specific oncogenes.* Oncogene, 1998. **16**(6): p. 737-46.

50. Chang, F., et al., *Signal transduction mediated by the Ras/Raf/MEK/ERK pathway from cytokine receptors to transcription factors: potential targeting for therapeutic intervention.* Leukemia, 2003. **17**(7): p. 1263-93.

51. Spencer, K.S., et al., *ErbB2 is necessary for induction of carcinoma cell invasion by ErbB family receptor tyrosine kinases.* J Cell Biol, 2000. **148**(2): p. 385-97.

52. 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.

53. Vial, E., E. Sahai, and C.J. Marshall, *ERK-MAPK signaling coordinately regulates activity of Rac1 and RhoA for tumor cell motility.* Cancer Cell, 2003. **4**(1): p. 67-79.

54. Schlessinger, J., *New roles for Src kinases in control of cell survival and angiogenesis.* Cell, 2000. **100**(3): p. 293-6.

55. Thomas, S.M. and J.S. Brugge, *Cellular functions regulated by Src family kinases.* Annu Rev Cell Dev Biol, 1997. **13**: p. 513-609.

56. Huang, C., et al., *The role of tyrosine phosphorylation of cortactin in the locomotion of endothelial cells.* J Biol Chem, 1998. **273**(40): p. 25770-6.

57. Lohi, O. and V.P. Lehto, *EAST, a novel EGF receptor substrate, associates with focal adhesions and actin fibers.* FEBS Lett, 1998. **436**(3): p. 419-23.

58. Provenzano, C., et al., *Eps8, a tyrosine kinase substrate, is recruited to the cell cortex and dynamic F-actin upon cytoskeleton remodeling.* Exp Cell Res, 1998. **242**(1): p. 186-200.

59. Shoelson, S.E., et al., *Specific phosphopeptide binding regulates a conformational change in the PI 3-kinase SH2 domain associated with enzyme activation.* Embo J, 1993. **12**(2): p. 795-802.

60. Band, V. and R. Sager, *Distinctive traits of normal and tumor-derived human mammary epithelial cells expressed in a medium that supports long-term growth of both cell types.* Proc Natl Acad Sci U S A, 1989. **86**(4): p. 1249-53.

61. Stampfer, M.R. and P. Yaswen, *Culture systems for study of human mammary epithelial cell proliferation, differentiation and transformation.* Cancer Surv, 1993. **18**: p. 7-34.

62. Stampfer, M.R., et al., *Gradual phenotypic conversion associated with immortalization of cultured human mammary epithelial cells.* Mol Biol Cell, 1997. **8**(12): p. 2391-405.

63. Asthagiri, A.R. and D.A. Lauffenburger, *Bioengineering models of cell signaling.* Annu Rev Biomed Eng, 2000. **2**: p. 31-53.

64. Hlavacek, W.S., et al., *Rules for modeling signal-transduction systems.* Sci STKE, 2006. **2006**(344): p. re6.

65. 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.

66. 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.

67. Neves, S.R. and R. Iyengar, *Modeling of signaling networks.* Bioessays, 2002. **24**(12): p. 1110-7.

68. Papin, J.A., et al., *Reconstruction of cellular signalling networks and analysis of their properties.* Nat Rev Mol Cell Biol, 2005. **6**(2): p. 99-111.

69. Kumar, N., et al., *Applying computational modeling to drug discovery and development.* Drug Discov Today, 2006. **11**(17-18): p. 806-11.

70. Takahashi, K., et al., *A multi-algorithm, multi-timescale method for cell simulation.* Bioinformatics, 2004. **20**(4): p. 538-46.

71. Alter, O., P.O. Brown, and D. Botstein, *Singular value decomposition for genome-wide expression data processing and modeling.* Proc Natl Acad Sci U S A, 2000. **97**(18): p. 10101-6.

72. Alter, O., P.O. Brown, and D. Botstein, *Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms.* Proc Natl Acad Sci U S A, 2003. **100**(6): p. 3351-6.

73. Strang, G., *Introduction to Linear Algebra*. 1993, Wellesley, MA: Wellesley-Cambridge Press.

74. Geladi, P. and B. Kowalski, *Partial Least Squares Regression: A Tutorial.* Analytica Chimica Acta, 1986. **185**: p. 1-17.

75. Hoskuldsson, A., *PLS Regression Methods.* Journal of Chemometrics, 1988. **2**: p. 211-228.

Chapter 2 Development and application of a high-throughput migration assay reveals HER2-mediated migration arising from increased persistence

2.1 Introduction to migration

The observation of cell movement was first recorded in 1863 by Rudolf Virchow, a scientist studying the leukocyte behavior [1]. Today we know that cell migration is critical to diverse biological processes such as embryogenesis, wound healing, inflammatory response, and angiogenesis [2-4]. In addition, migration plays a prominent role in cancer metastasis [5, 6]. The study of cell migration has been greatly aided by the recent development of novel experimental and computational tools. Nevertheless, since the movement of a cell through space is dependent of properties not only intrinsic to the cell but also those of the space it moves through, this complicated process is still poorly understood. For reviews of the biophysical and biochemical regulation of cell migration and the many factors regulating it, I refer the reader to [3, 4, 7, 8]. Here, I restrict the discussion to a brief introduction of the types of cell migration and the ways in which it is measured and analyzed. Although many recent advances in the field of cell migration have come from the study of cells in three dimensions, I do not touch on these studies, since the methods described herein are in 2D. However, mechanistic understanding in 2D has been shown to be relevant and useful for 3D migration, and it is hoped that the work presented here is no different [3]. Beyond the introduction,

Chapter 2.2 describes a novel high-throughput experimental approach to measuring migration and Chapter 2.3 presents the application of the procedure to understanding aspects of HER2-mediated migration.

The study of cell migration is typically divided into three categories defined by the chemical makeup of the space the cell is migrating within [9]. Due to the interplay between the extracellular space and the intracellular events governing migration, these three categories often have distinct signaling mechanisms associated with cell movement. The first of these categories is called chemokinesis, which describes the movement of a cell in the absence of any external orienting cues. In contrast, chemotaxis, a second category, describes directed cellular movement in the presence of a soluble chemical gradient. Finally, haptotaxis is directed cell movement in the presence of an anchored chemical gradient. The work described in this thesis and the computational models introduced in the next section deal with chemokinetic movement only. For further information about directed cell migration, the reader is referred to the following reviews [10, 11].

Once the category of migration to be studied is determined, there are still many factors that will affect movement within a particular cell type. Major factors include the extracellular matrix present (ECM, e.g. tissue culture plastic, fibronectin, collagen, laminin, or Matrigel) and the ligands present in solution [9]. Depending on the cell's own expression of various receptors that interact with the extracellular environment (e.g. integrins or growth factor receptors), cell migration will vary as ECM or ligand conditions change. In addition, the density of deposited cells on any given surface affects cell movement, with certain cell types preferring to move in collective groups and others

moving more effectively as single cells [1]. Many of these factors affecting cell migration are determined by the measurement assay used. Below, I briefly introduce some prominent assays used for the measurement of migration.

2.1.1 Measurement assays

The three most common ways to study migration *in vitro* are 2D cell tracking, scratch assays (wound healing assays) and filter assays [12]. 2D cell tracking typically involves the seeding of cells onto a surface such as tissue culture plastic, glass, or one modified with ECM, and then tracking cells via phase contrast microscopy or fluorescent microscopy (if the cells are appropriately labeled). Using this technique, one can measure cell movement at many time periods over a long time course, thereby generating data appropriate for the mathematical models discussed below. These assays typically involve the sparse seeding of cells, which can be a problem for cells that move as collective clusters or in sheets. Apart from the biological considerations, the 2D assay is also low throughput, not allowing for the study of many cell conditions in short periods of time.

The scratch assay begins with the growth of cells to confluence on a surface not typically modified by ECM. A 'wound' is then created by dragging a plastic object across a subsection of the cells to scrape them off. Cells migrate into the wound and wound area is recorded. Cell movement is subjected to a different set of steric hindrances in this assay as compared to the 2D assay, and single cell movement is not recorded. In

addition, wound area is typically not measured at comparable temporal resolution to 2D assays, yielding less data for computational analysis.

Filter assays typically involve the seeding of cells and then the observation at some later time of how many cells migrated through a filter membrane with pores of a defined diameter. This assay yields the least temporal information of the assays described in this section. The movement can be collective or single cell, but is a function of the cell's ability to move and squeeze through a pore. Filter assays have been developed for high-throughput analysis and are commonly used in the pharmaceutical industry for the screening of drugs.

2.1.2 Mathematical models of cell migration

The characterization of temporally resolved cell movement has been greatly aided by mathematical and computational analyses that have codified descriptions of this complicated process as well as identified salient mechanisms of movement that have been shown to be of biological importance. For chemokinetic movement, the most commonly used model for cell movement is called the persistent random walk (PRW) model [13]. The PRW model states that cell movement is a product of two processes, cell speed and the directionality of motion (cell persistence). As discussed in Chapter 2.3, speed and persistence have been shown to be under distinct biochemical control, and the parsing of these two behavior enhances our understanding of cell migration and how to manipulate it via pharmacological intervention.

The PRW model was initially proposed based on the observation that cell velocity (*v*) is correlated at short times but at long times becomes uncorrelated. This lead to the description of the velocity autocorrelation function as [13]:

$$G_v(\tau) \equiv \langle v(t+\tau) \bullet v(t) \rangle = S^2 e^{-\tau/P} \tag{1}$$

where S is root mean squared speed and P is the directional persistence time. The mean squared displacement $\langle d^2(t) \rangle$ can then be obtained from integrating equation 1 [13]:

$$\langle d^2(t) \rangle = \langle \vec{d}(t) \bullet \vec{d}(t) \rangle = 2 \int_0^t dt' \int_0^{t'} d\tau G_v(\tau) \tag{2}$$

Which, together with equation 1, yields:

$$\langle d^2(t) \rangle = S^2 P(t - P(1 - e^{-t/P})) \tag{3}$$

The net effect of speed and persistence on the movement of a cell can now be expressed in terms of diffusion. Setting the above equation equal to *4Dt* we have:

$$D = S^2 P / 2 \tag{4}$$

As an alternate to equations 2-4, we can also describe cell movement using a Markov Model. Specifically, we use an O-U process, descriptive of an auto-correlated random walk. The equation then is [14]:

$$dv(t) = -\beta v(t)dt + \sqrt{\alpha}dW(t) \quad (5)$$

where the first term on the RHS refers to the resistance to motion and the second term to random fluctuations. This description is often used when motivating the use of Monte Carlo (MC) models to simulate cell migration under varying conditions.

Using the above equations, estimates for *S*, *P,* and *D* can be obtained from experimental cell migration data via the fitting on equation 3 to displacement data [13]. We have included in Appendix 1 the MATLAB code used to do this. In addition, Chapters 2.3 and Chapter 6 demonstrate how these equations can be used to describe and characterize cell movement under varying biological conditions.

2.1.3 Connection to cancer and metastasis

As noted in Chapter 2.3, the migration work presented in this thesis was first motivated by the observation that HER2 overexpression correlates with increased metastasis in breast cancer patients. Metastasis, the process by which cells spread from a primary tumor to secondary sites, is comprised of many steps including separation of cells from the primary tumor, invasion of those cells through surrounding tissue and basement membranes, entry in the circulation, peritoneal, or lymphatic space, deposition at secondary site, and finally survival and proliferation after deposition [15]. Cell migration plays a key role in a number of these steps, most critically in the movement of a cell away from the primary tumor and toward the circulation as well as the relocation of the cell to its eventual secondary site. Indeed, cell migration has been shown to be a

required element in the metastasis of many cancers [5]. It stands to reason, then, that increased cell migration has been found to be a hallmark in highly invasive cancers, such neuroepithelial tumors [16]. Increased migration has been identified in a variety of cancers such as ovarian, breast, pancreatic, and prostate cancers [17, 18].

2.2 Development of a high-throughput assay

This section describes the development of a novel high-throughput assay to measure migration [19]. The assay combines the advantages of the above described scratch assays (e.g. high-throughput and allows cells to move collectively) with those from a 2D assay (temporally resolved information for single cells ideal for computational analysis). As described in Chapter 2.3 and Chapter 6, this assay can be used to quickly and accurately measure the effects of differential receptor, ligand, or inhibitor treatment on various cell lines. As compared with current technologies, it offers the unique combination of high-quality (e.g. many individual measurements) migration-specific information and assay speed.

2.2.1 Methods

Cell Culture

184A1 human mammary epithelial cells (HMEC's) were a kind gift from Martha Stampfer (Lawrence Berkeley Laboratory, Berkeley CA). Cells were maintained in DFCI-1 medium supplemented with 12.5 ng/ml EGF (Peprotech, Inc.). 184A1 HMEC clone 24H were a kind gift from Steve Wiley (Pacific Northwest National Laboratories,

Richland WA) and were maintained in DFCI-1 medium supplemented with 12.5 ng/ml EGF and 150 μg/ml of Geneticin (Gibco, Invitrogen Inc.). Serum-free DFCI-1 media is defined here as DFCI-1 without EGF, bovine pituitary extract, fetal bovine serum, and insulin.

Statistics

A Lilifors-Test was used to test for normality for speed, persistence, and random motility coefficient data. A Kolmogorov-Smirnov test was used to assess p-values at a 95% confidence interval for non-normal speed, persistence, and random motility coefficient data. P-values for wound healing comparisons were generated using two-sample t-test and ANOVA. All statistics were generated in MATLAB (Mathworks, Inc.).

Migration Assay

HMECs were seeded in 96-well tissue culture plastic plates (Packard View Plate Black, Ref. 6005225) at confluence (~500,000 cells/cm^2) and allowed to adhere for 4-6 hours. Media was then removed and cells were serum starved for 12-16 hours. Starved cells were treated for 30 minutes with 9μM 5-chloromethylfluorescein diacetate (CMFDA, Molecular Probes, Inc.) in serum-free media. CMFDA containing media was removed and cells were then treated with new serum-free media, serum-free media containing EGF (100 ng/ml), or serum-free media containing HRG-β1 (80 ng/ml, Sigma). A wound width ~650 μm was scraped in each well and cell movement imaged every 15 minutes for 12-15 hours using Cellomics KineticScan. For individual cell tracking in the monolayer, CMFDA labeled cells were mixed 1:20 with unlabelled cells and the mixture

was then seeded at confluence, serum starved, scraped and imaged as described above. Kinetics of wound closure were quantified using an in-house analysis software that calculated the wound area at each time point normalized by the initial wound area. The details of the computational code are given in Appendix 2. A 5-point time averaging algorithm was used to average wound closure in individual wells of equal treatment into a single trajectory at 30 minute intervals. Individual cell trajectories in monolayer were produced using Imaris (Bitplane, Inc.). Only cells that remained in frame throughout the entire experiment were considered. Each fluorescently labeled object was recognized as a single cell in a monolayer using the built-in spots function. Cell tracks of the fluorescent objects over time were generated with a built-in autoregressive motion algorithm [20]. Cell trajectories were then fit to the persistent random walk equation [13]. A schematic of this assay is shown below in Figure 2-1.

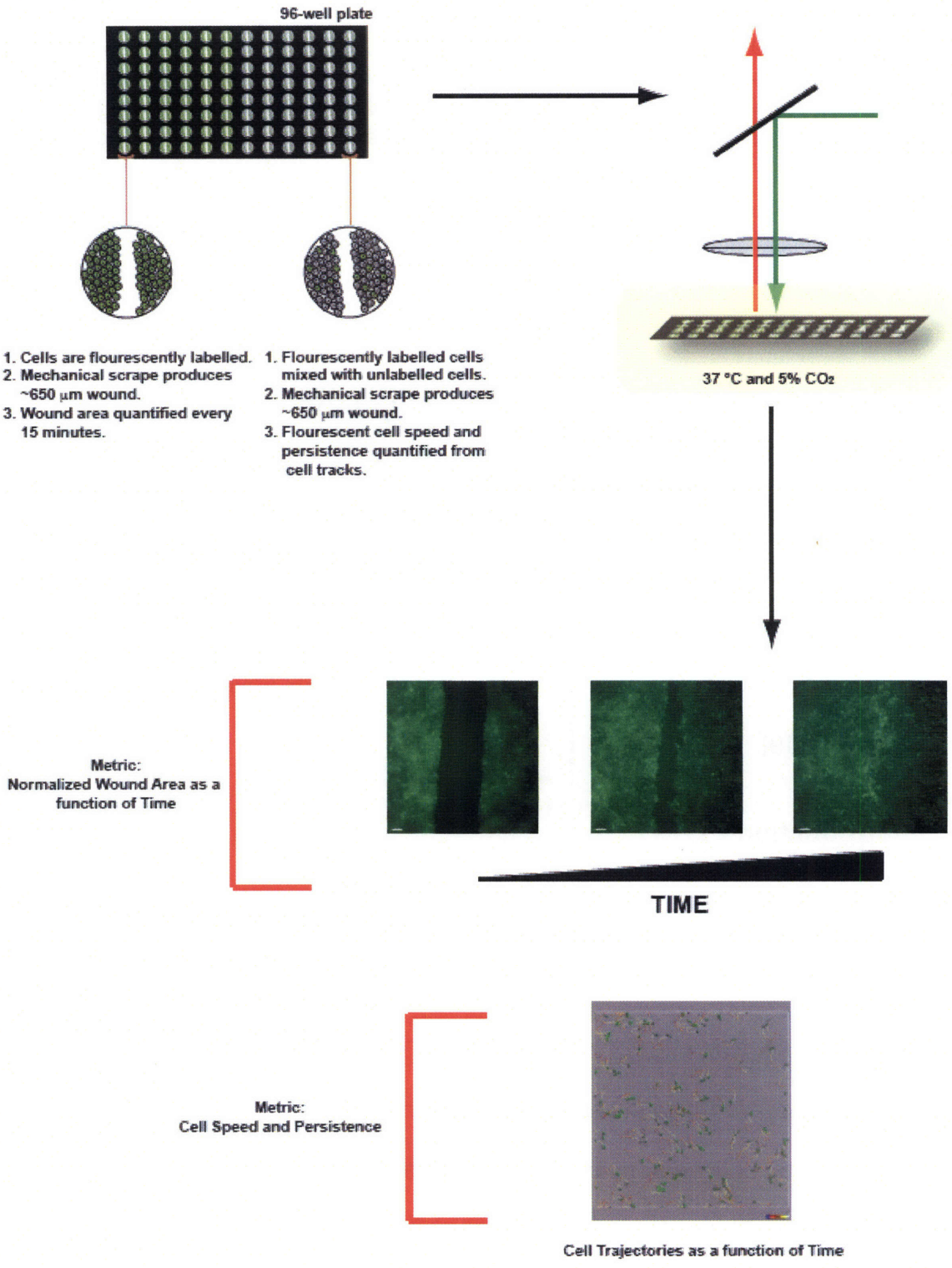


Figure 2-1: Schematic of high-throughput wound healing assay. Fluorescently labeled cells are either diluted with non-labeled cells or directly seeded and grown to confluence in a 96-well plate. Wounds ~650 μm in diameter are scraped in each well

and cell movement is then monitored using epi-fluorescent microscopy (Cellomics Kineticscan). The 96-well plate is kept at 37 °C and 5% CO_2 throughout the experiment. Movies are then exported and analyzed using MATLAB and Imaris-based software. Monolayer movement is quantified in terms of wound area and individual cell trajectories are defined in Imaris and then further analyzed for cell speed and persistence.

2.3 Measurements of HER2-mediated migration

This section describes the utilization of the assay described above to explore the effect of HER2 overexpression on cell speed and directional persistence. Observed consistent increases in persistence associated with HER2 overexpression indicate a prospective mechanism for invasiveness previously documented in HER2-overexpressing human breast tumors [19].

2.3.1 Introduction

Human epidermal growth factor receptor 2 (HER2) is overexpressed in 20% to 30% of breast cancers and correlates with poor prognosis and increased metastasis [21]. HER2 belongs to the ErbB or HER family of receptors (comprised of HER1/EGFR, HER2, HER3, and HER4) and can be activated through concentration-dependent homodimerization or ligand-driven heterodimerization. Epidermal growth factor (EGF) and heregulin (HRG), two ErbB family ligands implicated in cancer progression, bind HER1 and HER3, respectively, to induce the activation of HER2 through heterodimerization [22]. Because of HER2's role in breast cancer metastasis, a number

of groups have investigated the effect of HER2 expression on aspects of cell motility, demonstrating that activation by EGF, HRG, or homodimerization leads to increased invasion and motility in breast cancer cell lines [23-25]. In addition, these and other studies have implicated various downstream signaling molecules as effectors of HER2-increased motility. Primarily because many of these studies relied on invasion assays, however, there is little known about how HER2 overexpression affects cell migration parameters such as cell speed and persistence. Prior study of primary ductal breast carcinoma cells revealed that groups of cells tend to detach from primary tumor lesions and move away in a highly polarized and directionally persistent manner, indicating that the control of directional persistence may be distinct in highly motile breast cancer cells, such as those with HER2 overexpression [26]. In addition, increased directional persistence has been identified as a hallmark of cell migration in highly invasive tumors, such as neuroepithleal tumors [16]. Distinct signaling and biophysical mechanisms controlling directional persistence versus random motility have also been identified in recent studies [27-29]. Thus, a more in-depth study of HER2's effect on cell migration, speed, and persistence could potentially serve two purposes: 1) to connect HER2 overexpression with persistent movement shown to be important in cancer systems; and 2) to provide a context within which to understand previously identified HER2-associated downstream signals by linking them to pathways that regulate directionally persistent migration.

2.3.2 Wound healing results

We examined cell migration in a human mammary epithelial cell (HMEC) line. Two clones of the cell line, Parental (with 200,000 EGFR, 20,000 HER2, and 20,000 HER3) and 24H (200,000 EGFR, 600,000 HER2, and 30,000 HER3), were studied in the presence of EGF (100 ng/ml), HRG (80 ng/ml), or serum-free media. Cell migration was tracked using a high-throughput 96-well migration assay that we developed for the rapid screening of cell motility. The movement of epithelial monolayers and the motility of hundreds of individual cells in monolayers were rapidly screened, generating time-resolved population-level statistics for all treatment conditions during one 15-hour time course (Figure 2-1). This presages future assay application to drug-based screens designed to rapidly explore the biochemical basis of cell migration.

Increasing HER2 levels from 20,000 to 600,000 increased wound closure across all treatment conditions (Figures 2-2A-D). Absolute levels of wound closure differed depending on treatment condition, with EGF-treated cells being the most motile followed by HRG-treated cells under both low and high HER2 expression (Figure 2-2A). EGF-treated 24H cells exhibited the most rapid closure, completely sealing the wound in approximately 6 hours, and EGF treatment showed the most responsiveness to HER2 overexpression as measured by the difference in closure between parental and 24H cells at 15 hours. Interestingly, HER2 overexpression caused elevated closure even in the absence of ligand, presumably due to low levels of autocrine production or concentration-dependent homodimerization.

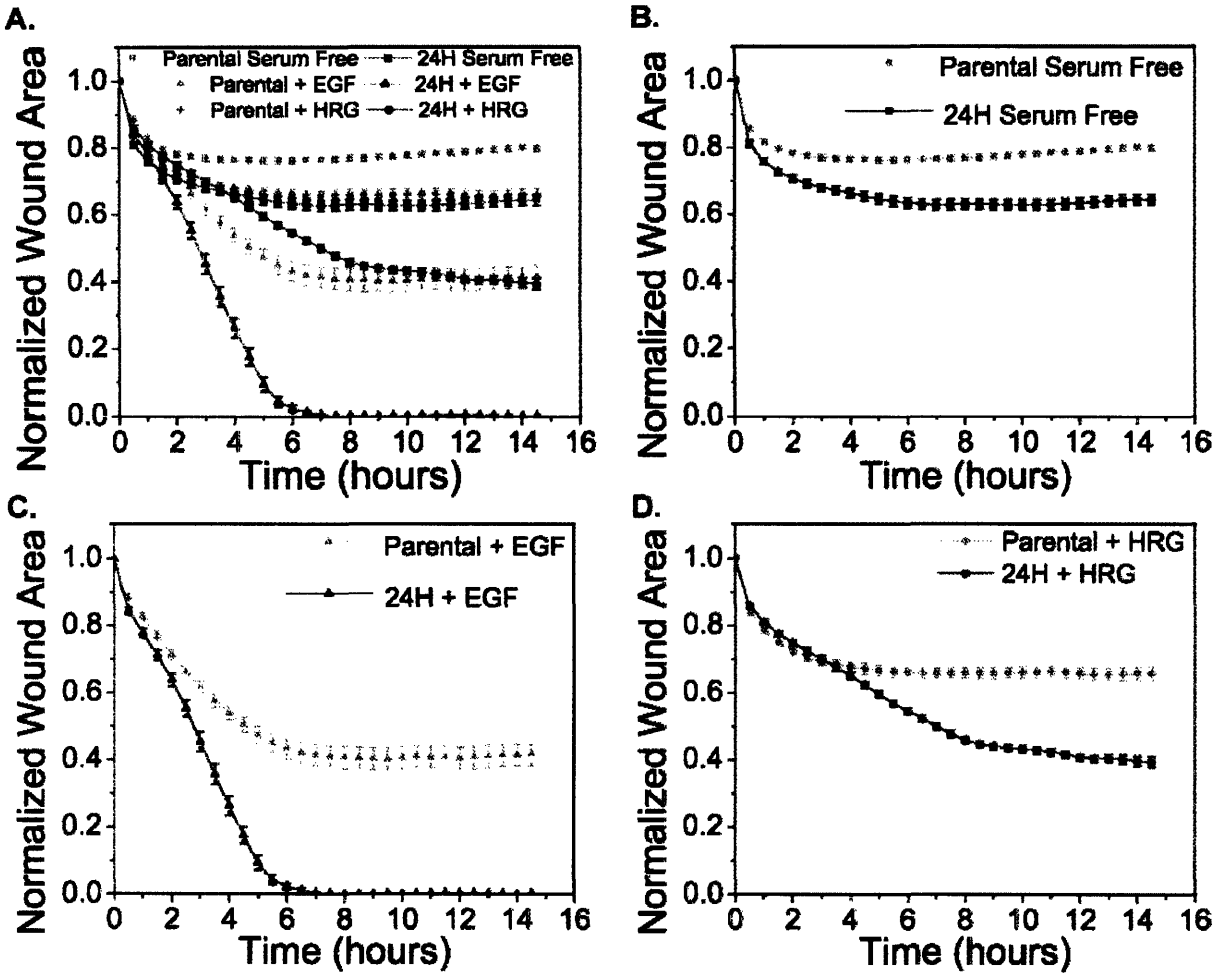


Figure 2-2: HER2 overexpression increases cell migration. Normalized wound area as a function of time for (A) parental and 24H cells under EGF (100 ng/ml), HRG (80 ng/ml), or serum-free treatment, (B) parental and 24H cells in serum-free conditions, (C) EGF-treated parental and 24H cells, and (D) HRG-treated parental or 24H cells. All time points shown ± SE after time averaging (see Supplementary Methods). For raw data time courses see Appendix 3, Figure A3-1.

Detailed inspection of wound-closure kinetics demonstrates time-dependent control of migration differing across treatment conditions. For instance, in the presence of HRG, HER2 overexpression increased wound closure significantly only after the first 3.5 hours, with both parental and 24H wounds measuring approximately 68% of the original area before separation ($p > 0.9$, means are the same). In contrast, serum-free

and EGF-treated 24H cells close more wound than similarly-treated parentals at 3.5 hours ($p < 0.01$), demonstrating that HRG exerts temporally distinct control of migration. Movement after 3.5 hours differentiates the HRG-treated cells, with 24H cells closing approximately 25% more normalized wound area than parentals at 15 hours. Another interesting kinetic trend is the cessation of wound closure that occurs at early times under certain conditions. For instance, HRG treated 24H cells close the wound throughout the duration of the experiment, but parental cells stop their movement after approximately 5 hours (Figure 2-2D). Interestingly, even after wound closure has stopped, HRG-treated cells continue to move in the monolayer, suggesting that cell movement perpendicular to the wound front is controlled separately from cell movement within the monolayer and parallel to the wound front (data not shown). EGF-treated parental cells exhibit similar behavior as do both parental and 24H cells in serum-free conditions. The time at which wound closure stops varies between approximately 2 to 6 hours depending on treatment conditions, but cell movement in the monolayer and parallel to the wound continues in all cases. These data are consistent with wound closure observations from other groups, and although the mechanistic basis of this 'stopping' behavior is not fully understood, it is clear that HER2 overexpression in the context of HRG or EGF treatment provides necessary information to overcome the signals governing the termination of wound closure before the wound is sealed [30]. Figure 2-2A also reveals that cells under all treatment conditions close similar wound areas at early times (21-29% wound area closed under all conditions at 1.5 hours). This suggests that early wound closure is primarily governed by mechanical induction and

associated start signals that operate independent of ligand/receptor conditions (Figure 2-3).

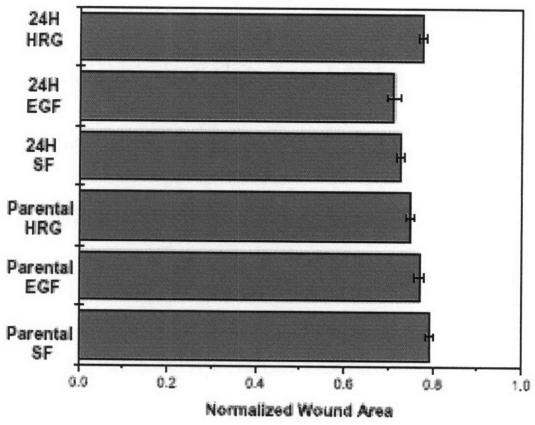


Figure 2-3: Early phase wound closure is similar for all cell treatments. Normalized wound area after 1.5 hours for 24H and parental cells under EGF (100 ng/ml), HRG (80 ng/ml), or serum-free conditions. Average wound area is reported ± SE.

2.3.3 Individual cell movement in monolayers

We considered whether the demonstrated differences in migration might arise from changes in cell speed, directional persistence, or both. We wondered further whether cells use these two migration 'levers' in a similar or different manner to increase migration under different ligand treatments. To explore these questions, we tracked the wound closure response of fluorescent cells diluted in unlabelled cells. Dilution allowed for more accurate cell tracking within the monolayer. We then calculated the mean-squared displacement for each cell trajectory and fit it to the following equation:

$$\left\langle d^2(t) \right\rangle = 2S^2P\left[t - P\left(1 - e^{-t/P}\right)\right] \qquad (1)$$

where S is cell speed and P is cell directional persistence [13]. Data demonstrated that although HER2 overexpression increased wound closure for all ligand conditions studied, it did not necessarily increase cell speed. Indeed, when HER2 was overexpressed, serum-free cell speed did not differ and HRG treated cells exhibited slightly decreased speed (Figure 2-4A). EGF-treated cells, however, increased cell speed with HER2 overexpression (Figure 2-4A). The magnitude of cell speed was similar between HRG treatment and serum-free conditions, but significantly higher for EGF-treated cells. In contrast to cell speed data, HER2 overexpression increased directional persistence across all treatments (Figure 2-4B). EGF treatment stimulated the highest degree of persistence, whereas HRG treatment induced the largest change between low and high HER2 conditions. Cell speed and persistence data together demonstrate that HER2 overexpression causes more rapid wound closure under EGF treatment due to increases in speed and persistence. However, in the absence of ligand or under HRG treatment, changes in closure rates are due primarily to increased cell persistence alone. HER2's influence on the effective diffusion of cells is found by equating equation 1 to 4Dt when t >> P, where D is the diffusion constant or random motility coefficient (as it is more often referred to in migration literature). Expressing D as:

$$D = S^2 P/2 \qquad (2)$$

we can evaluate the net effect of changes in speed and persistence on cell migration. Figure 2-4C shows that the random motility coefficient describes differences in the

magnitude of migration between conditions that are consistent with those observed using the wound closure assay (Figures 2-2A, 2-4C). Thus, depending on the ligand treatment, the cell regulates speed and persistence in qualitatively different ways to achieve HER2-mediated increases in migration as measured by wound closure or diffusion.

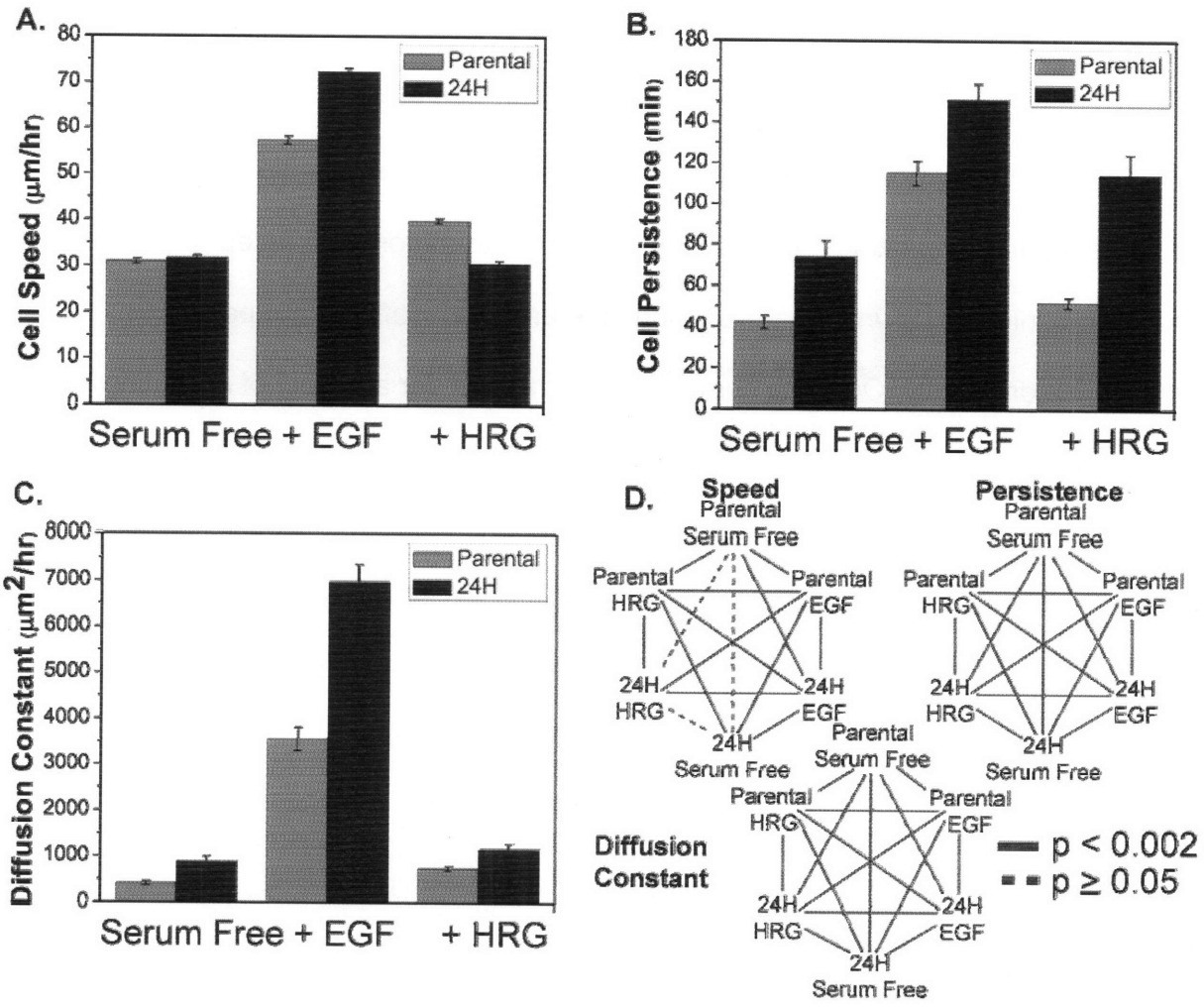


Figure 2-4: HER2 effects on cell speed and persistence. Cell migration of parental and 24H cells treated with EGF (100 ng/ml), HRG (80 ng/ml), or in serum-free conditions was recorded and quantified for (A) cell speed in μm/hr, (B) directional persistence in minutes, and (C) diffusion constant in $\mu m^2/hr$. (D) shows inequalities at 95% confidence as evaluated using Kolmogorov-Smirnov test for non-normal

distributions. All data reported as mean ± SE. Number of cells per condition equals 153-196. (See Appendix 3, Figures A-2-A2-4 for raw data)

2.3.4 Conclusions

The biological regulators of directional movement are not fully understood, but recent work has implicated Rac1, microtubules, and myosin IIb as possible candidates [16, 27, 28]. Whatever the downstream regulators, our data indicate that HER2 overexpression affects cellular components involved with directional movement independently from those responsible for increased cell speed. We speculate that differential phosphorylation under the various ligand treatment and receptor levels considered initiates downstream signaling differences that affect cell speed and persistence differently. Our data show, however, that under a variety of activating conditions, HER2 overexpression increases cell persistence, perhaps indicating the existence of a redundant phospho-site or signaling mechanism responsible for increasing persistence. This strong connection between HER2 and directionally-persistent migration may, in turn, be an important facet to HER2's documented ability to increase invasion and metastasis in human breast cancer cells.

2.4 References

1. Friedl, P., Y. Hegerfeldt, and M. Tusch, *Collective cell migration in morphogenesis and cancer.* Int J Dev Biol, 2004. **48**(5-6): p. 441-9.
2. Feldner, J.C. and B.H. Brandt, *Cancer cell motility--on the road from c-erbB-2 receptor steered signaling to actin reorganization.* Exp Cell Res, 2002. **272**(2): p. 93-108.
3. Lauffenburger, D.A. and A.F. Horwitz, *Cell migration: a physically integrated molecular process.* Cell, 1996. **84**(3): p. 359-69.
4. Ridley, A.J., et al., *Cell migration: integrating signals from front to back.* Science, 2003. **302**(5651): p. 1704-9.
5. Wang, W., et al., *Tumor cells caught in the act of invading: their strategy for enhanced cell motility.* Trends Cell Biol, 2005. **15**(3): p. 138-45.
6. Yamaguchi, H., J. Wyckoff, and J. Condeelis, *Cell migration in tumors.* Curr Opin Cell Biol, 2005. **17**(5): p. 559-64.
7. Maheshwari, G. and D.A. Lauffenburger, *Deconstructing (and reconstructing) cell migration.* Microsc Res Tech, 1998. **43**(5): p. 358-68.
8. Zamir, E. and B. Geiger, *Molecular complexity and dynamics of cell-matrix adhesions.* J Cell Sci, 2001. **114**(Pt 20): p. 3583-90.
9. Zetter, B.R. and S.E. Brightman, *Cell motility and the extracellular matrix.* Curr Opin Cell Biol, 1990. **2**(5): p. 850-6.
10. Schneider, I.C. and J.M. Haugh, *Mechanisms of gradient sensing and chemotaxis: conserved pathways, diverse regulation.* Cell Cycle, 2006. **5**(11): p. 1130-4.

11. Van Haastert, P.J. and P.N. Devreotes, *Chemotaxis: signalling the way forward.* Nat Rev Mol Cell Biol, 2004. **5**(8): p. 626-34.

12. Entschladen, F., et al., *Analysis methods of human cell migration.* Exp Cell Res, 2005. **307**(2): p. 418-26.

13. Dickinson, R. and R. Tranquillo, *Optimal estimation of cell-movement indexes from the statistical-analysis of cell tracking data.* AICHE Journal, 1993. **39**(12): p. 1995-2010.

14. Stokes, C.L., D.A. Lauffenburger, and S.K. Williams, *Migration of individual microvessel endothelial cells: stochastic model and parameter measurement.* J Cell Sci, 1991. **99 (Pt 2)**: p. 419-30.

15. Hunter, K.W., *Host genetics and tumour metastasis.* Br J Cancer, 2004. **90**(4): p. 752-5.

16. Deisboeck, T.S., T. Demuth, and Y. Mansury, *Correlating velocity patterns with spatial dynamics in glioma cell migration.* Acta Biotheor, 2005. **53**(3): p. 181-90.

17. Friedl, P. and K. Wolf, *Tumour-cell invasion and migration: diversity and escape mechanisms.* Nat Rev Cancer, 2003. **3**(5): p. 362-74.

18. Pienta, K.J. and R. Loberg, *The "emigration, migration, and immigration"of prostate cancer.* Clin Prostate Cancer, 2005. **4**(1): p. 24-30.

19. Kumar, N., et al., *A high-throughput migration assay reveals HER2-mediated cell migration arising from increased directional persistence.* Biophys J, 2006. **91**(4): p. L32-4.

20. Veenman, C., M. Reinders, and E. Backer, *Resolving motion correspondence for densely moving points.* IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 2001. **23**(1): p. 54-72.

21. Hynes, N.E. and D.F. Stern, *The biology of erbB-2/neu/HER-2 and its role in cancer.* Biochim Biophys Acta, 1994. **1198**(2-3): p. 165-84.

22. Harari, D. and Y. Yarden, *Molecular mechanisms underlying ErbB2/HER2 action in breast cancer.* Oncogene, 2000. **19**(53): p. 6102-14.

23. Brandt, B.H., et al., *c-erbB-2/EGFR as dominant heterodimerization partners determine a motogenic phenotype in human breast cancer cells.* Faseb J, 1999. **13**(14): p. 1939-49.

24. Dittmar, T., et al., *Induction of cancer cell migration by epidermal growth factor is initiated by specific phosphorylation of tyrosine 1248 of c-erbB-2 receptor via EGFR.* Faseb J, 2002. **16**(13): p. 1823-5.

25. Spencer, K.S., et al., *ErbB2 is necessary for induction of carcinoma cell invasion by ErbB family receptor tyrosine kinases.* J Cell Biol, 2000. **148**(2): p. 385-97.

26. Friedl, P., et al., *Migration of coordinated cell clusters in mesenchymal and epithelial cancer explants in vitro.* Cancer Res, 1995. **55**(20): p. 4557-60.

27. Lo, C.M., et al., *Nonmuscle myosin IIb is involved in the guidance of fibroblast migration.* Mol Biol Cell, 2004. **15**(3): p. 982-9.

28. Pankov, R., et al., *A Rac switch regulates random versus directionally persistent cell migration.* J Cell Biol, 2005. **170**(5): p. 793-802.

29. Verkhovsky, A.B., T.M. Svitkina, and G.G. Borisy, *Self-polarization and directional motility of cytoplasm.* Curr Biol, 1999. **9**(1): p. 11-20.

30. Jacinto, A., A. Martinez-Arias, and P. Martin, *Mechanisms of epithelial fusion and repair.* Nat Cell Biol, 2001. **3**(5): p. E117-23.

Chapter 3 Signal state measurement techniques as applied to HMEC signaling

In this chapter, I discuss the various signal measurement technologies employed in this thesis. Each method is introduced along with pertinent HMEC results, and assay-specific issues are addressed. Finally, a comparison between the technologies is presented to motivate the use of various assays in Chapters 4-7.

3.1 Kinase activity assay

3.1.1. Methods

The methodological details for the high-throughput microtiter kinase assay used to study kinase activity in the HMEC cells has been described elsewhere [1]. Figure 3-1 shows a generalized schematic of the process. Briefly, antibody raised against the kinase of interest is incubated with Reacti-Bind™ protein A or G coated microtiter wells overnight at 4°C. Following washing, wells are incubated with whole cell lysate overnight at 4°C. After washing, each well is suspended in kinase buffer containing cold ATP and [γ-^{32}P]ATP, and the reaction is initiated with the appropriate substrate. After incubation at 37°C, the reaction is terminated with 75 mM H_3PO_4 or 20 mM EDTA. The reaction contents are filtered through a 96-well phosphocellulose filter plate using the Multiscreen® system (Millipore, Bedford, MA). After extensive washing, the ^{32}P incorporation is quantified by liquid scintillation. For each set of experiments, blank wells containing only lysis buffer are included as negative controls. This procedure has

been validated for the following molecules of interest: Akt, JNK, ERK, and IκB in HT-29 cells [1].

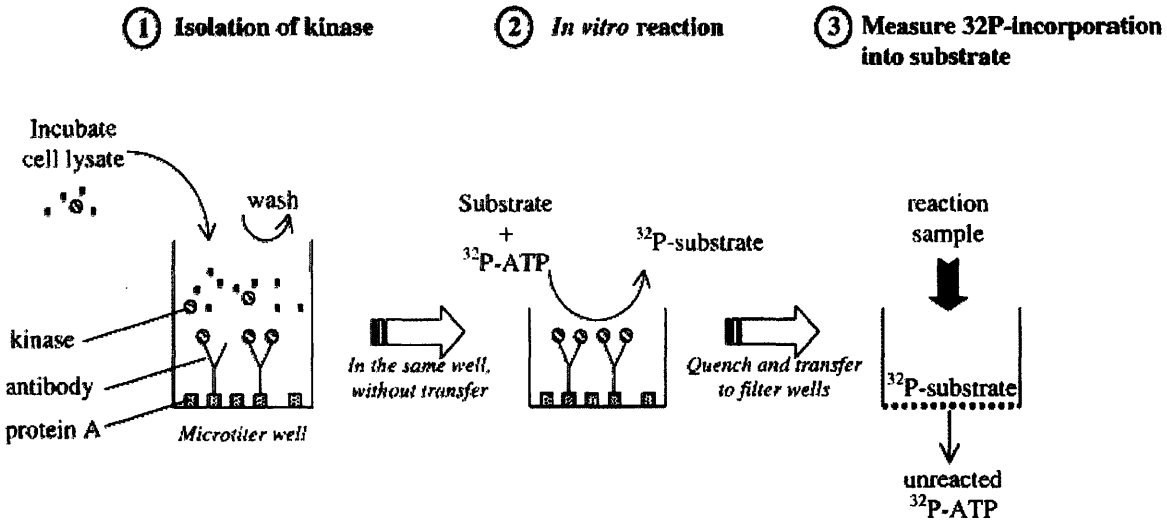


Figure 3-1: Schematic of kinase activity assay [1].

3.1.2 HMEC results

To assess the applicability of the kinase activity assay in the HMEC cell system, we tested each assay for linearity. We found that, using the previously published methodology, we could achieve linearity for both the Erk and Jnk assays (Figure 3-2). We were not, however, able to obtain a linearity curve for Akt or IκB kinases (Figure 3-2). Interestingly, IκB activity was very high as compared to HT-29 cells, indicating a either a high concentration of IκB or an increase in the percent of IκB activity in HMEC cells as compared to HT-29 cells. For both Akt and IκB, we tried to vary the following parameters to establish linearity: incubation time of lysates with antibody, antibody

coating concentration, the ratio of radioactive ATP to non-radioactive ATP, and reaction time.

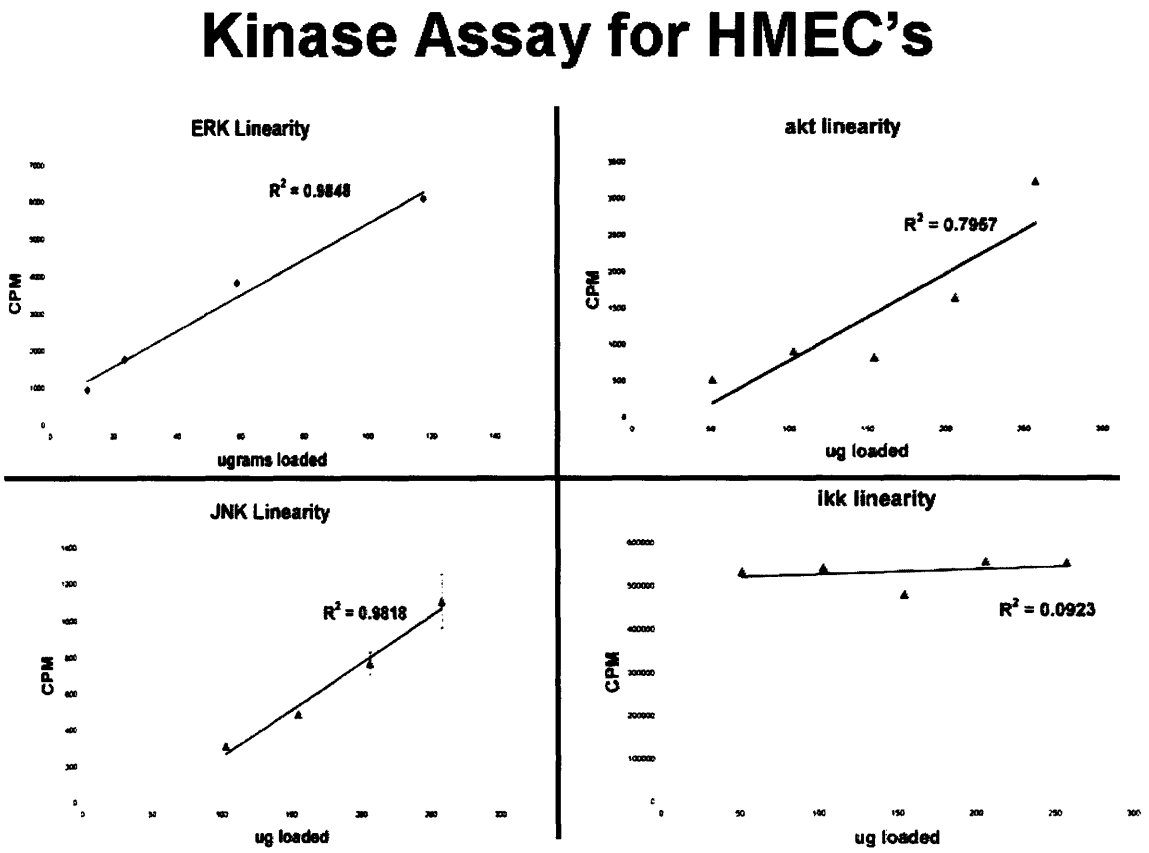


Figure 3-2: Linearity test for HMEC kinase assay. Serially diluted lysate was used to test measurement linearity for Erk, Akt, Jnk, and IKK kinase activity assays.

3.1.3 Key issues

Kinase activity assays directly measure kinase activity. If it is the kinase's ability to phosphorylate substrate that one is interested in, then activity measurements are a more direct way to asses that. The other methods mentioned here rely on measurement of phosphorylated kinase, which typically, but not always, predicts kinase activity (see

Chapter 7). However, a disadvantage of the activity assay is the use of peptide substrate which substitutes for the variety of natural substrates found in the cell. Kinases may have differential abilities to phosphorylate naturally occurring substrates in the cell, due to issues such as structure or concentration, therefore rendering the activity assay results potentially misleading. In addition, the washes that occur prior to the kinase activity reaction remove most proteins bound to the kinase. These proteins may serve to regulate kinase activity in the cell, and their absence may be another reason for misleading results.

3.2 Western blot

3.2.1. Methods

To quantify phosphorylation levels, lysate is resuspended in sample buffer [100 mM DTT, 2% SDS, 10% glycerol, 0.01% bromophenol blue, 62.5 mM Tris-HCl (pH 6.8)]. Gel electrophoreses (10% polyacrylamide gel) is followed by transfer to polyvinylidene difluoride membranes (Biorad). Membranes are blocked with 5% nonfat milk or 5% bovine serum albumin in 20 mM Tris-HCl (pH 7.5), 137 mM NaCl, and 0.1% Tween-20. Membranes are then probed with an antibody specific to the kinase of interest. The membranes are then probed with horseradish-peroxidase-conjugated anti-rabbit secondary antibody (Amersham Pharmacia Biotech) and visualized by enhanced chemiluminescence (Amersham Pharmacia Biotech) on a Kodak Image Station (Perkin Elmer). Densitometry was performed using molecular imaging software (Kodak). Linearity for each antibody can be established using serial dilutions of lysate.

3.2.2 Key issues

Perhaps the most critical issue for western blotting in quantitative work is the large error associated with quantification of blots. Typical blot-to-blot heterogeneity, due to factors such as transfer efficiency, tends to exaggerate measurement error and, compared to the other technologies covered here, does not lend itself to quantitative work. Chapter 7, entitled "Quantitative analysis of Akt phosphorylation and activity in response to EGF and insulin treatment," deals with quantitative western blotting and highlights some of the problems surrounding the approach. Although it may have some weaknesses for quantitative studies, western blotting still remains an extremely useful technology for the validation of quantitative trends obtained with other technologies. The ability to ascertain the specificity of antibodies and the consistency of the western blot assay make it essential for both preliminary studies and validation studies.

3.3 Antibody array

3.3.1 Methods

The methodology associated with the antibody array we used (BioSource Mercator Array Assay, BioSource, Inc.) can be found elsewhere [2]. A schematic for the procedure is shown below in Figure 3-3. Briefly, lysates were deposited on nitrocellulose pads arrayed on a glass slide. Each nitrocellulose pad had spots corresponding to monoclonal capture antibodies against kinases of interest. In

necessary for the antibody chip assay but not used in western blotting. Nevertheless, given that the antibody array data did not agree with western blots or intuition about the signaling system, we were not able to use the dataset for further studies. We do not, however, discount the future possibility of the array's use for signaling studies and are encouraged by its internal consistency.

3.4 Immunocytochemistry

3.4.1 Methods

We have published the methodology for the immunocytochemistry technique elsewhere [3] and a full description is available in Chapter 6. A schematic of the assay is shown below in Figure 3-6.

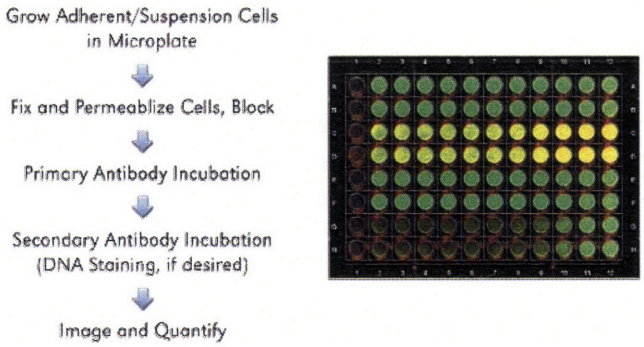


Figure 3-6: Schematic of immunocytochemical assay [5].

3.4.2 HMEC results

Extensive characterization of HMEC cell signaling using the immunocytochemical assay has been published [3] and is shown in Chapter 6.

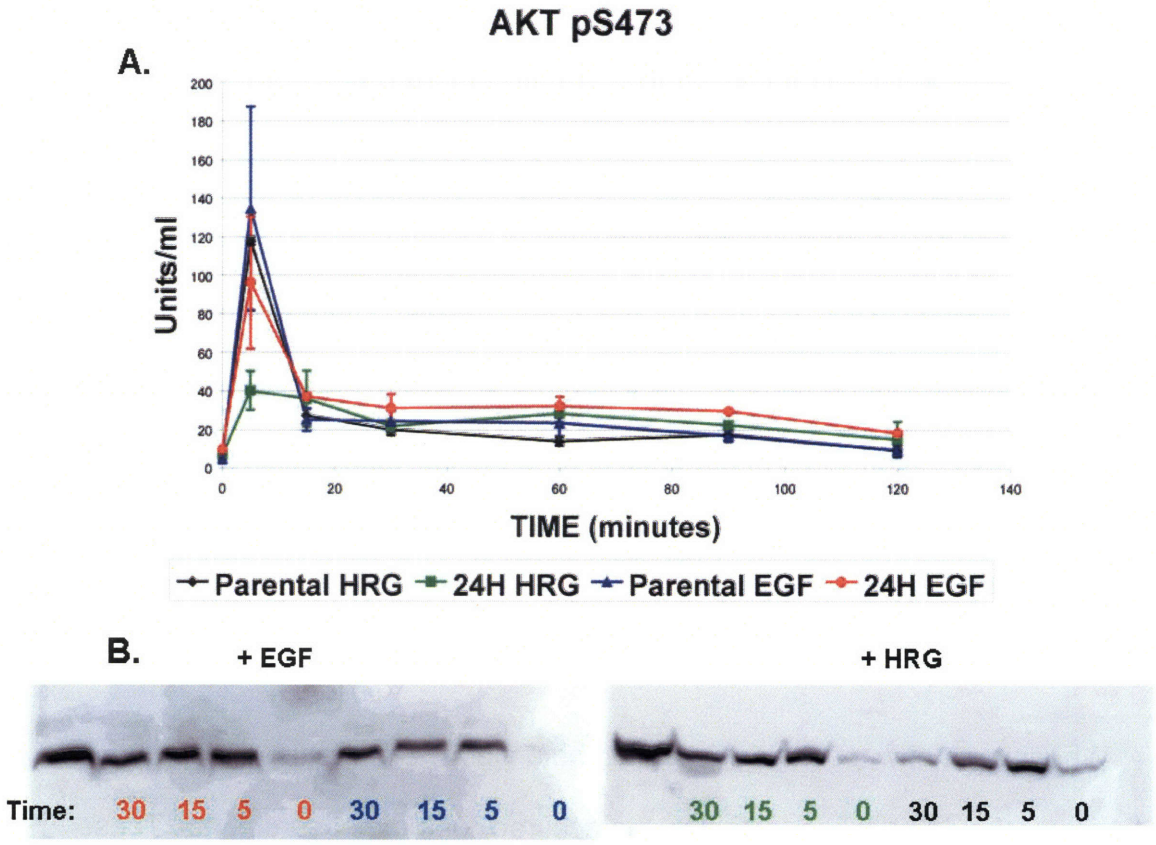


Figure 3-4: Akt phosphorylation in parental and 24H cells. Measurements from the Mercator Antibody Chip (A) are compared with those from western blots on the same lysates (B). Cells were stimulated with either HRG (80 ng/ml) or EGF (100 ng/ml). All data shown ±SEM.

We found measured transient activation in both parental and 24H cells stimulated with EGF or HRG. HRG-stimulated 24H cells seemed to be stimulated to a smaller degree than the rest of the cells, but excepting the early time point (5 minutes), there were no significant differences between cell types. Similar measurements of EGFR, Fak, Src, and Paxillin are shown in Appendix 4. EGFR measurements revealed activation with EGF stimulation, but not with HRG stimulation, which agrees with HRG's role as a promoter of HER3-HER2 heterodimers and data we have acquired using

immunocytochemistry and mass spectrometry [3, 4]. In addition, Y1068 phosphorylation is more sustained in 24H cells, agreeing with our previous data as well. Src phosphorylation measurements, although very noisy, revealed a transient spike in parental cell stimulated with EGF that was not observed in 24H cells. 24H cells sustained with EGF, however, exhibited late-phase activation as compared to parental cells. Fak phosphorylation revealed a transient but significant increase in HRG-stimulated 24H cells, but no phosphorylation in under any other cell treatment. There were not statistically significant differences among the cell data acquired for paxillin phosphorylation.

3.3.3 Key issues and validation

Whenever a new technology such as the antibody array is used, appropriate careful steps should be taken to ensure the accuracy of the measurements. These steps should include validation within the assay and validation of the assay as compared with other assays. In the case of the antibody array, validation within the assay revealed a very consistent measurement technique. Calibration curves generated from serially diluted recombinant amounts of phosphorylated kinases were extremely consistent on different chips and on different days (see Figure 3-5 below). In addition, the standard error of the mean from seven different measurements of lysis buffer on seven different chips was less than 5% of the average measured value, indicating high chip-to-chip reproducibility.

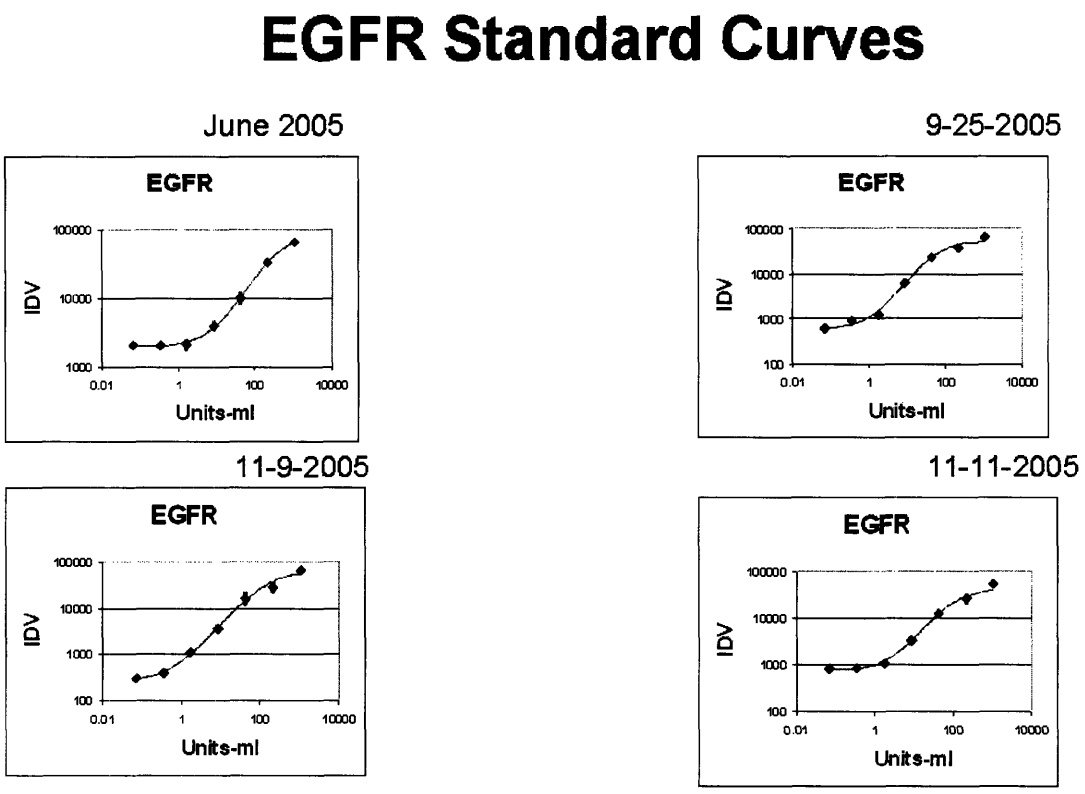


Figure 3-5: Chip-to-chip error for an antibody array. Integrated intensity (IDV) plotted against differing amounts of recombinant phosphorylated EGFR (in ml) for four different antibody chips assayed on four different days.

The reproducibility of the assay, however, could not be verified for certain kinases by other measurement techniques. For instance, the Akt data gathered did not agree with measurements we made using immunocytochemistry (see Chapter 6) or western blot data measuring phosphorylation on the exact same lysates used in the antibody chip data (Figure 3-4A). A similar disagreement between western blot data and Fak phosphorylation data was observed (Appendix 4, Figure 2). Thus, validation attempts through a separate assay failed to corroborate all antibody chip data. One reason for the noted discrepancies could be due to the immunoprecipitation step

necessary for the antibody chip assay but not used in western blotting. Nevertheless, given that the antibody array data did not agree with western blots or intuition about the signaling system, we were not able to use the dataset for further studies. We do not, however, discount the future possibility of the array's use for signaling studies and are encouraged by its internal consistency.

3.4 Immunocytochemistry

3.4.1 Methods

We have published the methodology for the immunocytochemistry technique elsewhere [3] and a full description is available in Chapter 6. A schematic of the assay is shown below in Figure 3-6.

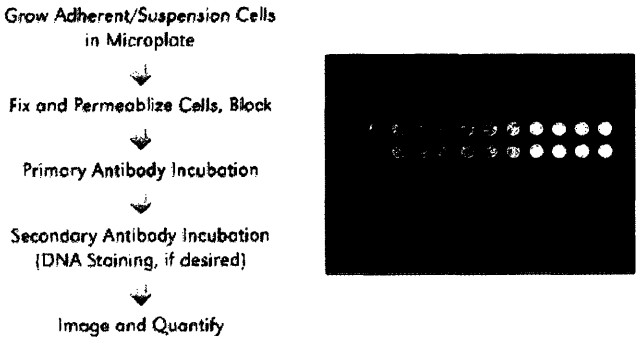


Figure 3-6: Schematic of immunocytochemical assay [5].

3.4.2 HMEC results

Extensive characterization of HMEC cell signaling using the immunocytochemical assay has been published [3] and is shown in Chapter 6.

3.4.3 Key issues and validation

The immunocytochemical assay was subjected to internal validation and validated against other signaling assays. An obvious limitation of the technique is that any antibody having non-specific binding properties will not be useful in such an assay.

All monoclonal antibodies used were validated using western blot analysis for specificity using lysates from parental and 24H cell lines after stimulation with full serum (see Chapter 6).

To measure plate-to-plate accuracy, we compared normalized serum-free values for 24H cells run on different plates but on the same day. These values should be equal assuming no plate-to-plate variation. We observed a very tight correlation between the values for 24H, i.e. the +HRG 24H and the +EGF 24H conditions. The variation was less than 10% of the total signal in 19 of 20 measured comparisons.

3.5 Mass spectrometry

3.5.1 Methods

The specifics of the mass spectrometry approach that used in this thesis can be found in Chapters 4, 5 and in a paper by Zhang et al [6]. Here, I introduce the basic setup and methodology employed. An approach utilizing a stable-isotope-tagged amine-reactive reagent (iTRAQ) labeling system was employed to quantify levels of phosphorylated tyrosine at four different time points (Figure 3-7). Cell lysates corresponding to various time points were digested, fractionated, and each time point was incubated with a unique iTRAQ label. These labels are isobaric but differ with respect to the location of atomic isotopes, allowing for their separable resolution.

Peptide fragments were immunoprecipitated using anti-phosphotyrosine antibody, further enriched, and then loaded onto a C18 column. Ions were generated using electrospray ionization (ESI) and a quadrupole time-of-flight (Q-TOF) instrument characterized the mass-to-charge (m/z) characteristics of the peptides. Further fragmentation of high-intensity peptide ions generated MS/MS spectra that was then sorted against a protein database for peptide identification. This LC-MS/MS approach coupled with the use of iTRAQ enabled the measurement of hundreds of tyrosine phosphorylation sites under various cellular conditions.

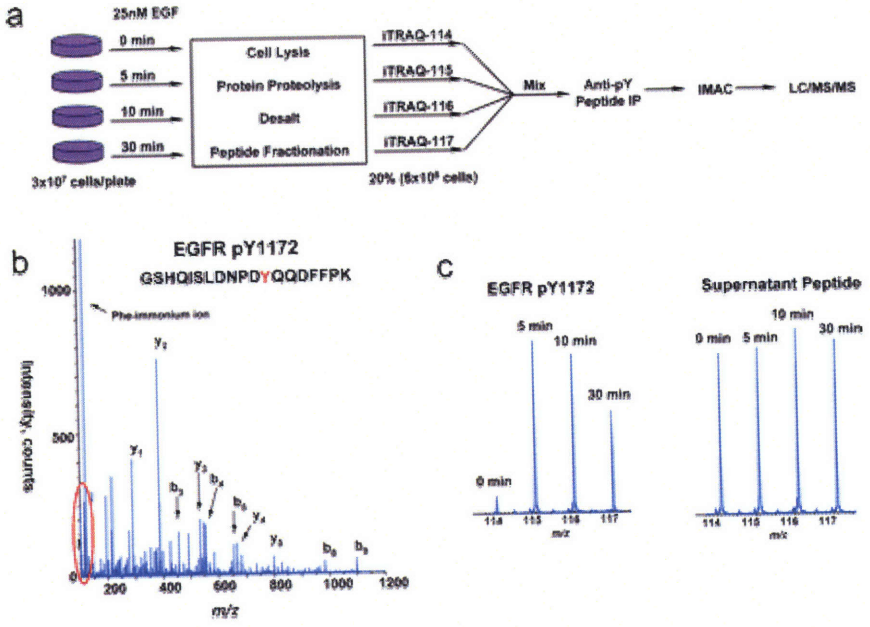


Figure 3-7: Mass spectrometry methodology [6]. (A) Cells at four different time points are lysed and eventually processed to produce iTRAQ-labeled phosphotyrosine peptide ion fragments. (B) MS/MS spectra are analyzed and reveal the presence of specific phosphotyrosine sites. (C) Analysis and appropriate normalization of the iTRAQ-associated spectra enable the temporal quantification of phosphorylation on a specific site.

3.5.2 HMEC results

The application of the above techniques to HMEC cell lines resulted in the quantification of 62 phosphorylation sites in both parental and 24H cell stimulated with EGF or HRG. Measurements were taken at 0, 5, 10, and 30 minutes. These data are presented in Chapters 4 and 5, and in Wolf-Yadlin et al. and Kumar et al. [3, 4].

3.5.3 Key issues

Although the signal state coverage enabled by the mass spectrometry technique is large compared with the other techniques described here, many measured peptides were removed from our final dataset due to the fact that they were not observed in all of the MS runs. This lack of identification cannot simply be explained by a lower quantity of phosphorylated peptide; rather, there is also the possibility that the peptide was accidentally 'skipped' by the MS instrumentation during one run. This problem can be resolved by moving from information dependent acquisition (IDA, as used here in our study) to multiple reaction monitoring (MRM), which allows for higher measurement reproducibility across samples (personal communication, Alejandro Wolf-Yadlin).

Manual validation of individual peptide ion fragments is also necessary for the validation of the final MS dataset. This validation procedure is time-consuming (taking on the order of a month for our HMEC dataset), thus limiting the number of cellular conditions that can be analyzed per unit time. Additional validation was performed on the dataset by means of standard deviation analysis, with those samples exhibiting more that 15% deviation being dropped from further analysis. These validation procedures, while time consuming, ensure the accuracy of the final dataset and the resultant quantitative models. The significant overlap between the phosphorylation

profiles measured by the other techniques mentioned and those measured in the mass spectrometer suggest a high degree of accuracy in the MS datasets procured here.

3.6 Conclusions

This chapter provides an introduction to the signaling technologies we employed to interrogate the signal state of HMEC cells. There are inherent trade-offs from moving between these technologies, and specifically, we were concerned with performance in the following areas: 1) time from lysis to analyzed data, 2) quantitative accuracy, 3) signal state coverage. The first area of interest, time from lysis to analyzed data, is a function of the speed of the assay and the length of the corresponding analysis. Here, both the immunocytochemical and antibody array assays offer a significant advantage; we were able to rapidly acquire and analyze data for many different cell conditions using these assays. The mass spectrometry technique is the slowest of the technologies in this respect. As mentioned earlier, the western blot technique is most likely the weakest the of the performed assays in terms of quantitative accuracy. The other assays typically yield 10-20% error as a percentage of the measurement when studied using different biological replicates, which we feel is an accurate portrayal of the biological noise in the system. An important note with regard to the quantitative accuracy of the immunocytochemical assay is that linearity is difficult to explicitly establish. Serial dilution of lysate is obviously not possible in this format, and a dilution of cell density would most likely lead to signaling differences due to density changes, thus confounding linearity experiments. The final category of interest is signal state coverage, and here the mass spectrometry technique is clearly the superior technology. We were able to

measure 62 different phosphorylation sites using this technique, and recent advances should increase that number to in excess of 150. The coverage available using the immunocytochemical assay is limited due to the lack of specific monoclonal antibodies for this purpose. The antibody array's coverage is most likely limited due to this factor and the problem of cross-binding as you increase the number of targets on any one spot on the slide. Finally, although western blotting can offer large signal coverage in theory, in practice a single investigator would find it difficult to take advantage of the many targetable sites, although services do now exist for multi-dimensional profiling via western blot.

References

1. Janes, K.A., et al., *A High-throughput Quantitative Multiplex Kinase Assay for Monitoring Information Flow in Signaling Networks: Application to Sepsis-Apoptosis.* Mol Cell Proteomics, 2003. **2**(7): p. 463-73.
2. Biosource, *www.biosource.com*.
3. Kumar, N., et al., *Modeling HER2 effects on cell behavior from mass spectrometry phosphotyrosine data.* PLoS Computational Biology, 2006. **in press**.
4. Wolf-Yadlin, A., et al., *Effects of HER2 overexpression on cell signaling networks governing proliferation and migration.* Mol Syst Biol, 2006. **2**: p. 54.
5. Li-cor, *www.licor.com/bio*.
6. Zhang, Y., et al., *Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules.* Mol Cell Proteomics, 2005. **4**(9): p. 1240-50.

Chapter 4 Defining signal-network control of HER2-mediated cell behavior

In this Chapter, we demonstrate how to generate, combine, and interpret quantitative cell signal and cell behavior data in a HER2 overexpressing HMEC system.

4.1 Introduction

HER2, also known as Erb-B2, belongs to the epidermal growth factor receptor (EGFR) family of highly regulated receptor tyrosine kinases (RTKs) composed of human epidermal growth factor receptor 1,2,3 and 4 (EGFR, HER2, HER3, HER4). These receptors are differentially bound by 11 ligands, although so far no ligand has been found to bind HER2. Mutation and dysregulation of EGFR family members has been correlated with cancer development and progression [1, 2], and overexpression of HER2 has been found in association with a variety of tumor types [3]. The connection between HER2 and cancer appears to be especially important in breast cancer [4], with current estimates of prevalence on the order of 25-30% of patients including metastases as well as primary tumors [5] and notable correlation with poor prognosis [6]. For instance, one recent review has concluded that HER2/neu gene amplification or HER2 protein overexpression can predict breast cancer outcome for 90% of the studies and 92% of the patients in 81 studies including 27,161 patients [7].

Although HER2 appears to have no intrinsic ligand-binding capability, it can interact reversibly with ligand activated EGFR or HER3 to form active heterodimers that perturb and often enhance the downstream signals that govern cell proliferation and migration

[3]. At high concentrations, HER2 is also thought to spontaneously form signaling-competent homodimers [8]. In the case of HER2-EGFR heterodimers, signaling proceeds through phosphorylation events initiated from either of the activated receptors. In contrast, HER2-HER3 heterodimer signaling proceeds only through the activated HER2 receptor, since HER3 does not exhibit kinase activity itself. HER2 is considered responsible for at least the vast majority of downstream signals induced by HER3-ligand binding [9], although there is some evidence of HER3-EGFR signaling heterodimers [10]. Activation of HER2-HER3 heterodimers, generally considered to be the most potent signaling pair of EGFR family dimers [11], is tightly regulated [6], and HER2 over-expression has been found to mediate enhanced signaling caused by heregulin (HRG) stimulation [12]. Various studies have shown that HRG-stimulated HER3 activation in breast cancer can induce anti-estrogen resistance, tumor progression, invasion and metastasis[13-15]. Taken in sum, the high correlation between EGFR family member dysregulation and cancer progression highlights the need for mechanistic understanding of the underlying cellular networks, both for improved basic knowledge of cancer and to find new and more effective drug targets.

Signal generation and resultant cellular behavior in receptor tyrosine kinase-(RTK-) initiated cascades is achieved through a highly coordinated series of events. Receptors are first activated, typically through ligand binding, which results in receptor auto-phosphorylation and leads to the dynamic phosphorylation/de-phosphorylation of a variety of proteins. The dynamic state of protein phosphorylation ultimately controls cellular response, and a given protein may influence more than one response as it becomes phosphorylated at different sites or interacts with different proteins depending

on timing and placement within the cell. Thus, to understand cellular signaling in context of cell regulation it is helpful to quantify phosphorylation dynamics of regulatory sites and their consequent association with downstream cell functions. Our work presented here offers, to our knowledge, the most detailed explanation of how signaling network phosphorylation, in the context of HER2 overexpression, governs cellular behavior such as proliferation and migration across a diverse set of ligand stimulation conditions.

To identify important EGF and HRG induced protein phosphorylation events that control cell migration and proliferation in the context of HER2 overexpression, we utilized a mass spectrometry approach that is capable of simultaneously quantifying the temporal dynamics of a large number of specific tyrosine phosphorylation sites under a given condition [16]. Here we extend this methodology to measure temporal dynamics of protein tyrosine phosphorylation sites following EGFR or HER3 stimulation in the presence and absence of HER2 overexpression. Protein phosphorylation sites are clustered based on similarity of dynamic response to different stimulation conditions across 4 time points for all 4 conditions (self-organizing map), revealing co-regulated phosphorylation sites and providing potential functionality for several novel phosphorylation sites. In addition to mass spectrometric analysis of protein phosphorylation, biological response data (cell migration and proliferation) was quantified for each of the four stimulation conditions. To identify signals that regulate downstream biological response to a given stimulus, computational methods were used to correlate biological responses to quantitative phosphoproteomics data.

Phosphorylation sites which correlate strongly with proliferation or migration were identified and may be targeted in future studies to selectively inhibit a given response.

4.2 Methods and materials

4.2.1 Cell culture and stimulation

184A1 HMECs (human mammary epithelial cells [17]) (HMEC Parental) were a kind gift from Martha Stampfer (Lawrence Berkeley Laboratory, Berkeley CA) via Steve Wiley (Pacific Northwest National Laboratory, Richland WA and were maintained in DFCI-1 medium supplemented with 12.5 ng/ml EGF, as in [18]. 184A1 HMECs clone 24H (human mammary epithelial cells over-expressing HER2 30 fold [18]) (HMEC 24H) were a kind gift from Steve Wiley (Pacific Northwest National Laboratories, Richland WA) and were maintained in DFCI-1 medium supplemented with 12.5 ng/ml EGF and 150 μg/ml of Geneticin. Cells were washed with PBS and incubated for 12 hours in serum free media (DFCI-1 without EGF, bovine pituitary extract, or fetal bovine serum) after 80% confluence was reached in 15 cm plates (~ 2×10^7 cells). Synchronized cells were washed with PBS after removal of media. Cells were then stimulated with 100 ng/ml EGF or 80 ng/ml HRG in serum free media for 5, 10 or 30 minutes, or left untreated with serum free media for 5 min as control.

4.2.2 Mass spectrometry lysate preparation and analysis

Cell Lysis, Protein Digestion, Peptide Fractionation and iTRAQ Labeling

Cells were lysed with 8M urea / 1mM sodium orthovanadate after EGF or HRG stimulation. Proteins were digested with trypsin after DTT reduction and iodoacetamide alkylation. Tryptic peptides were desalted and fractionated on a C18 Sep-Pak Plus cartridge (Waters), and the 25% MeCN fraction was divided into 10 equivalent aliquots which were lyophilized to dryness. One aliquot of sample from each condition was labeled with one tube of iTRAQ reagent (Applied Biosystems, CA) (following manufacturer's directions). Samples labeled with four different isoforms of the iTRAQ reagent were combined, dried completely, and saved at -80°C. This process was repeated to generate five duplicate sets of samples: four time-course samples (0, 5, 10, 30 min) with 100 ng/ml EGF or 80 ng/ml HRG stimulation in either HMEC or 24H cells, and one 5 min mix sample that consisted of the samples stimulated for 5 min for each of the stimulation conditions in the order of HMEC/HRG, HMEC/EGF, 24H/HRG, 24H/EGF.

Peptide IP.

10 µg of protein G Plus-agarose beads (Calbiochem) were incubated with 3.5 µg of each anti-phosphotyrosine antibody (PY99(Santa Cruz), 4G10(Upstate) and pTyr100 (Cell Signaling Technology)) in 200 µl of IP buffer for 8 hr at 4°C. The beads were rinsed once with 400 µl of IP buffer at 4°C. iTRAQ labeled sample were dissolved in 150 µl of IP buffer (100 mM Tris, 100 mM NaCl, 1% NP-40, pH 7.4) and 300ul of water. After pH was adjusted to 7.4 with 0.5 M Tris buffer pH 8.5, the sample was mixed with protein G Plus-agarose beads, and was incubated overnight at 4°C. The protein G Plus-agarose beads were spun down for 5 minutes at 6000 rpm and the supernatant was

separated and saved. Antibody-bound beads were washed once with 400 µl IP buffer for 10 minutes and twice with rinse buffer (100 mM Tris, 100 mM NaCl, pH 7.4) for 5 minutes at 4°C. Phosphotyrosine containing peptides were eluted from antibody with 60 µl of 100 mM glycine pH 2.5 for 30 min at room temperature.

IMAC and Mass Spec.

Phosphopeptide enrichment on IMAC was performed as described [16]. Peptides retained on the IMAC column were eluted to a C_{18} capillary precolumn (100 µm ID, 360 µm OD) with 50 µl of 250 mM Na_2HPO_4, pH 8.0. After 10 min rinse (0.1% HOAc), the precolumn was connected to an analytical capillary column with an integrated electrospray tip (~1 µm orifice). Peptides were eluted using a 100 min gradient with solvent A (H_2O/HOAc, 99/1 vol/vol) and B (H_2O/MeCN/HOAc, 29/70/1 vol/vol): 10 min from 0% to 15% B, 75 min from 15% to 40% B, 15 min from 40% to 70% B. Eluted peptides were directly electrosprayed into a quadrupole time-of-flight mass spectrometer (QSTAR XL Pro, Applied Biosystems). MS/MS spectra of four or five most intense peaks with 2 to 5 charges in the MS scan were automatically acquired in information-dependent acquisition with previously selected peaks excluded for 25 seconds.

Phosphopeptide Sequencing, Data Clustering and Analysis.

MS/MS spectra were extracted and searched against human protein database (NCBI) using ProQuant (Applied Biosystems) as described. An interrogator database was generated by predigesting the human protein database with trypsin and allowing one miscleavage and up to six modifications on a single peptide (phosphotyrosine ≤ 2, phosphoserine ≤ 1, phosphothreonine ≤ 1, iTRAQ-lysine ≤ 4, and iTRAQ-tyrosine ≤ 4).

Mass tolerance was set to 2.3 amu for precursor ions and 0.15 amu for fragment ions and grouping of MS/MS spectra from different cycles was set to zero. Phosphotyrosine-containing peptides identified by ProQuant were manually validated. ProQuant quantification results were corrected by removing contaminant signals near iTRAQ tag peaks. Data was further corrected with values generated from the peak areas of non-phosphorylated peptides to account for possible variations in the starting amounts of sample for each time point. All the data from each analysis was normalized by the 115 peak area values, which correspond to the 5 min sample in each time-course analysis or HMEC-EGF sample in the four-condition 5 min mix analysis.

4.2.3 Hierarchical clustering

After normalization to the 5 minute HMEC-EGF sample, hierarchical clustering of the four-condition 5min Mix analysis was done using Spot Fire™ using the Euclidean distance and unweighted average settings.

4.2.4 Self-Organizing Maps

The idea behind the SOM is to non-linearly transform high-dimensional input data into a lower dimensional display that consists of several map units. Each map unit consists of a reference vector whose dimension is the same as the dimension of an input pattern [19]. The clustering is done into two steps. In the training phase the SOM algorithm computes distance between an input pattern and all reference vectors. The

map unit consisting of reference vector that correlates the best with the input pattern is declared as the winner map unit and is updated with maximal strength, while topologically close map units are updated with gradually decreasing strength. This process is executed using all 62 peptide profiles and repeated a few times leading to self-organization. After the training phase each peptide is assigned to the map unit whose reference vector is the most similar to the input pattern.

The parameters for the SOM used in this study were as follows. Topology of the map was chosen to be sheet, distance measure was correlation and the number of map units was chosen to be $5 \cdot \sqrt{p}$, where p is the number of input vectors as suggested in [20]. We used the batch learning algorithm and the neighborhood function was chosen to be Gaussian with the parameters given in [20]. The SOM analysis was performed in MATLAB with the publicly available SOM Toolbox [20].

We used the U-matrix (unified-distance matrix) method to identify a group of map-units that comprehend a cluster [21]. The U-matrix illustrates Euclidean distances between the map units with colors. Adjacent map units colored with blue shades constitute a cluster (valleys) and the clusters are separated red and yellow colors that represent high distance between two map units (mountains). For each cluster we computed statistical significance with the permutation test based method given in [22] as follows. First we computed correlation distances for all combinations of the peptide profiles in a cluster. If the two profiles correlate perfectly, their distance is zero and perfect negative correlation results in the distance value two. Then we computed the mean of these pair-wise comparisons. This was followed by choosing randomly the same number of profiles as there are in the original cluster and computing all

combinations of the pair-wise correlation distances. For example, if a cluster consisted of 18 peptides, we randomly chose 18 peptides from the set of 62 peptides and computed the mean of 153 correlation distance values (18 choose 2), which was compared to the mean of the peptide correlations in the original cluster. If the mean distance of a randomly chosen set is less or equal than the original we added one to counter. For each cluster we created 5,000 samples and the final *p*-value is computed by dividing the counter by 5,000. Large values suggest the original cluster may be due to chance.

4.2.5 Elisa for ErbB3 receptor quantification

Reagents for the ErbB3 ELISA were purchased from R&D Systems as a Duoset DY348 (capture antibody, biotinylated detection antibody, recombinant human Erb3 and streptavidin-HRP). A black 96-well Nunc MaxiSorp plate was coated with 50 □L of capture antibody (4 □g/mL in PBS) overnight at room temperature and washed three times with PBS, 0.05%Tween-20 on a Bio-Tek plate washer. The unreacted surface of the plate was blocked with 300 □L of 2% BSA in PBS for 1 hr at room temperature. After washing, 50 □L of each sample or the recombinant standard were added to the appropriate wells. Cell line lysates were mixed 1:1 with 2% BSA and added to the plate for a 2 hr incubation at 4°C. Both standards and samples were done in duplicate. After washing, 60 □L of detection antibody (2 □g/mL diluted in 2 %BSA, PBS, 0.1% Tween-20) was added to all the wells and incubated for 1 hr at room temperature. The plate was washed and 60 □L of Streptavidin-HRP (diluted 1:200 in 2 %BSA, PBS, 0.1%

Tween-20) was added for 30 min at room temperature. After a final wash, 60 □L of chemiluminescent substrate (SuperSignal ELISA Pico from Pierce) was added to all wells and the plate was read immediately on a Fusion plate reader (Perkin Elmer) with a 1 sec reading time per well. The data was exported to Excel, replicates were averaged and the standard dilutions were fit to a linear curve.

4.2.6 Proliferation assay

Human mammary epithelial cells were plated on 24 well tissue culture plastic plates (~1 X 10^4 cells/cm^2) and grown for 24 hours to ~60-70% confluence. The medium was then removed and cells were serum-starved as previously described for 12-16 hours. Starved cells were then treated with new serum-free media, serum-free media containing EGF (100 ng/ml), or serum-free media containing HRG (80 ng/ml). The cells were grown for 15 hours at which time [^{3}H] thymidine (10 μCi/ml) was added. After 10 hours, the cells were washed with cold PBS and treated with trichloroacetic acid (5% w/v) at 4°C. The resulting precipitate was dissolved in NaOH (.5 N) and quantified using liquid scintillation counting.

4.2.7 Migration assay

HMECs were seeded in 96-well tissue culture plastic plates (Packard View Plate Black, Ref. 6005225) at confluence (~50,000 cells/cm^2) and allowed to adhere for 4-6 hours. Media was then removed and cells were serum starved for 12-16 hours as

previously described. Starved cells were treated with 5-chloromethylfluorescein diacetate (CMFDA, 9 μM in serum-free media) for 30 minutes. CMFDA containing media was removed and cells were then treated with new serum-free media, serum-free media containing EGF (100 ng/ml), or serum-free media containing HRG (80 ng/ml). A wound width ~700 μm was scraped in each well and cell movement imaged every 15 minutes for 12 hours using Cellomics KineticScan. Kinetics of wound closure were quantified using an in-house analysis software that calculated the wound area at each time point normalized by the initial wound area. A time averaging algorithm was used to average wound closure in four wells of equal treatment into a single trajectory at 30 minute intervals. The trajectories shown here have been additionally normalized to their 2 hour time points for the purpose of comparing treatments. The wound closure data used in the described PLSR model is the linearly fitted slope (2 - 5.5 hours) of each trajectory shown.

4.2.8 Partial least squares regression

Computational analysis was performed using the SIMCA-P 10.0 (Umetrics) software suite as detailed elsewhere [23, 24] The software uses the non-linear iterative partial least-squares algorithm to perform decompositions and regressions. The model was evaluated for goodness of fit (R^2Y), goodness of prediction (Q^2), and was validated against over-fitting via response permutation (response matrix was randomly permuted 50 times and Q^2 values were obtained for each run). All matrices were mean centered and unit variance scaled (z-score normalized) prior to partial least squares analysis.

Goodness of prediction (Q^2) was evaluated using a leave-one-out cross validation approach [25]. Briefly, cross validation is performed by omitting an observation from the model development and then using the model to predict the Y-block values for the withheld observation. The procedure is repeated until every observation has been kept out exactly once. The prediction error sum of squares (PRESS) is then calculated as follows:

$$PRESS = \sum_i \sum_m Y_{i.m}^{measured} - Y_{i.m}^{predicted} \quad \textbf{(1)}$$

Q^2 is then calculated as:

$$Q^2 = 1.0 - \prod_{a=1}^{A} (PRESS / SS)_a \quad \textbf{(2)}$$

Where *a* refers to each individual principal component in the model and *SS* is the sum of squares explained by principal component *a*.

The variable importance for projection (VIP) value for each variable (k) was calculated according to the following formula:

$$VIP_k = \sqrt{\frac{K_T \sum_{a=1}^{A} w_{a,k}^2 SS_a}{\sum_{a=1}^{A} SS_a}} \quad \textbf{(3)}$$

Where K_T is the total number of variables and the rest of the variables are as described above.

4.3 Results and discussion

4.3.1 Phosphotyrosine mass spectrometry

We have undertaken a quantitative investigation of the effects of HER2-overexpression in a human mammary epithelial cell (HMEC) line (184A1 [17]). We compared the parental line (denoted "P"), which exhibits approximately 200,000 EGFR per cell, 20,000 HER2 per cell, and 20,000 HER3 per cell, to a stable retrovirally-transduced clone (denoted "24H") that expresses HER2 at 600,000 per cell while maintaining constant EGFR levels and increased HER3 levels to about 30,000 per cell. We have previously performed combined experimental and modeling studies on this cell line enabling quantitative estimation of the various receptor dimer species under various treatment conditions [18, 26]. From our previous work, we expect that 100 ng/ml EGF treatment should result in high levels of EGFR homodimers in the P HMECs (~80,000), but lower numbers of EGFR-HER2 heterodimers (~20,000). By comparison, treating 24H HMECs with this same concentration of EGF should drive a large increase in heterodimers (to ~150,000) and a significant decrease in EGFR homodimers (~10,000). 80 ng/ml HRG treatment should yield ~15,000 HER2-HER3 heterodimers in the P HMECs, and in the 24H HMECs this number should increase to ~25,000. Interestingly, under both ligand treatments, 24H HMECs are expected to have large numbers of HER2 homodimers (>~200,000), some of which may be activated through basal auto-phosphorylation. Here, we explore the consequences of these changing dimerization states on intracellular signaling and the subsequent control of cell proliferation and migration. We have acquired mass spectrometry data describing the temporal dynamics (0, 5, 10, and 30 minute stimulation) of tyrosine phosphorylation in the 184A1 cell system under four conditions: P HMECs stimulated with EGF, P HMECs stimulated

with HRG, 24H cells HMECs stimulated with EGF, and 24H HMECs stimulated with HRG (see Figure 4-1).

As a result of these analyses, 332 phosphorylated peptides from 175 proteins were identified, including 289 singly (tyrosine) phosphorylated peptides, 42 doubly phosphorylated peptides (21 tyrosine/tyrosine, 18 serine/tyrosine, and 3 threonine/tyrosine), and one triply phosphorylated peptide (tyrosine/tyrosine/tyrosine) (full dataset available in supplemental material). A total of 20 phosphorylation sites were identified on EGFR, HER2 and HER3, including 9 tyrosine and 2 serine sites on EGFR, 8 tyrosine phosphorylation sites on HER2, and 1 tyrosine phosphorylation site on HER3 (Figure 1, B-D). Of the 20 phosphorylation sites on EGFR family members, Y1114 on EGFR and Y1005 and Y1127 on HER2 represent novel sites which have not been previously described in the literature, although synthetic phosphopeptides containing EGFR Y1114 and HER2 Y1005 have been shown to bind SHC [27] and mutation of EGFR Y1114 has been shown to block SOCS recruitment to the receptor, regulating STAT activation [28]. Downstream of the receptors, quantitative data was obtained for 36 phosphorylation sites on 15 different proteins in the EGFR canonical signaling pathway (as defined by Yarden et al. [2]). Coverage of the cell adhesion/migration pathway included quantitative information on 41 phosphorylation sites distributed along 16 proteins, including 9 tyrosine phosphorylation sites on δ-catenin.

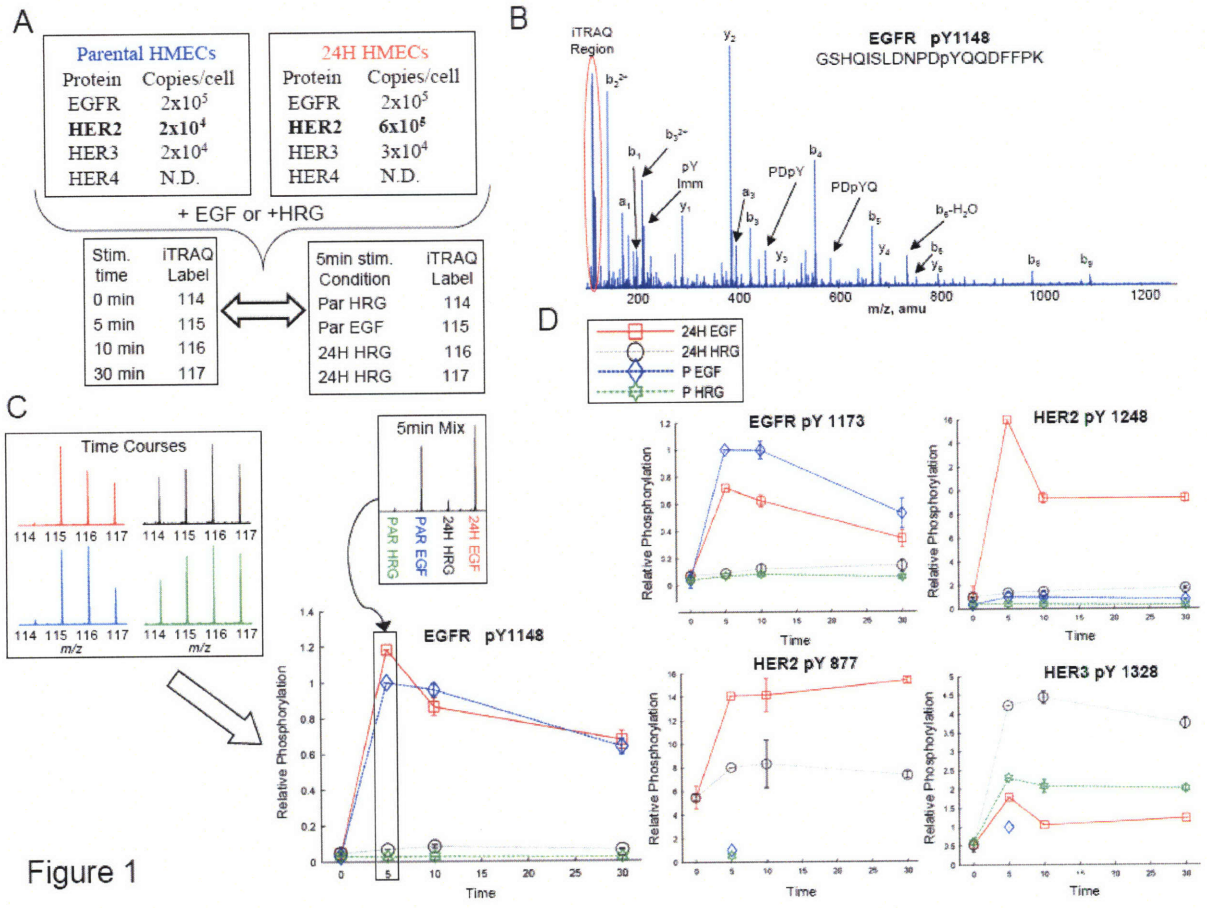


Figure 1

Figure 4-1: Data acquisition and quantification scheme with example data. (A) 184A1 parental HMEC cells and 24H HMEC cells were stimulated with either EGF or HRG for 0, 5, 10 or 30 minutes. For each stimulation condition, following cell lysis and tryptic digestion, peptides from each time point were labeled with iTRAQ, mixed, and analyzed by anti-phosphotyrosine peptide immunoprecipitation and IMAC-LC-MS/MS, producing temporal tyrosine phosphorylation profiles for hundreds of peptides. To normalize temporal profiles from each condition, tryptic peptides from the 5 minute stimulation time point for each condition were labeled with iTRAQ, mixed, and analyzed by anti-phosphotyrosine peptide immunoprecipitation and IMAC-LC-MS/MS. (B) For each peptide, y- and b-type ions in the MS/MS spectrum provided sequence identification and site assignment, while the iTRAQ marker ion region (highlighted in the red oval) provided quantification. (C) iTRAQ marker ion profiles representing temporal phosphorylation profiles for EGFR pY1148 under all four conditions (top left) as well as the relative amount of phosphorylation following a 5 minute stimulation for each condition (top right). The final plot of temporal response under multiple stimulation conditions (bottom) was generated by multiplying each temporal phosphorylation profile by the value obtained from the 5 minute mix quantification data for the particular stimulation condition. (D) Multiple tyrosine phosphorylation sites were identified and quantified from EGFR family member proteins. 4 representative phosphorylation sites are shown here, with response plots across temporal and conditional space. As can be

seen by comparing HER2 pY877 and pY1248, different sites on a given protein are often differentially regulated.

One of the goals in this study was to provide a more comprehensive definition of EGFR family signaling networks, including novel proteins or phosphorylation sites which may regulate differential cellular response to exogenous stimuli. To enable identification of novel phosphorylation sites, mass spectrometry analyses were performed in information dependent acquisition mode (automated selection of the most abundant species in a given full scan mass spectrum for MS/MS analysis), rather than targeting specific peptides and known phosphorylation sites for quantification. This data-dependent mode of operation was successful at identifying novel proteins and phosphorylation sites, as 122 of the 322 phosphorylation sites have not previously been described in the literature. Unfortunately, the use of automated ion selection to discover novel phosphorylation sites often precluded selection of low-abundance precursor ions for MS/MS fragmentation, and therefore temporal phosphorylation profiles were not obtained for all conditions for many peptides. In fact, 234 of the 322 sites in this work were detected and quantified in multiple analyses, but only 68 were quantified for all 4 stimulation conditions and the 5 minute comparison. This core group of 68 phosphopeptides, which include many of the key signaling nodes in the network, was then used for further computational analysis to obtain a mechanistic understanding of the effects of HER-2 overexpression on cellular signaling networks and corresponding biological response to growth factor stimulation.

4.3.2 Self-Organizing Maps define temporal and conditionally related clusters of phosphorylation sites

In order to identify clusters of tyrosine-phosphorylated peptides exhibiting similar temporal dynamics, as well as to globally visualize the high-dimensional information we have obtained, we used the self-organizing map (SOM) algorithm [19]. The SOM is a versatile clustering algorithm that transforms high-dimensional data into lower dimensional display, in a non-linear manner. Here we use the unified-distance matrix (or, U-matrix) approach, which allows robust identification of clusters, and the component plane representation which facilitates comparison of peptide phosphorylation response to exogenous stimuli [19, 21]. Instead of taking a single map-unit as one cluster, we use U-matrix to identify groups of map-units that together comprehend a cluster. The U-matrix illustrates the mean distances between neighboring map units after the SOM training phase. These distances are color-coded so that close proximity of two map-units is colored with bluish colors, while shades of yellow and red denote dissimilarity. Clusters can be identified as continuous bluish regions (valleys) surrounded by yellow or red "mountains."

We applied this SOM algorithm to 62 "core" phosphorylated peptides for which temporal profiles were generated under all conditions (six of the "core" phosphorylated peptides had minimal response across all conditions and time points and were removed from this analysis). Data were normalized with the z-score method across all time-points and conditions. From the U-matrix in Figure 4-2 four clusters are readily identifiable. Statistical significances for these clusters were computed with a permutation test based method (Hautaniemi et al., 2003) and the clusters were found to be statistically significant ($p < 0.05$). The entire SOM display including the component planes is given in the supplementary material.

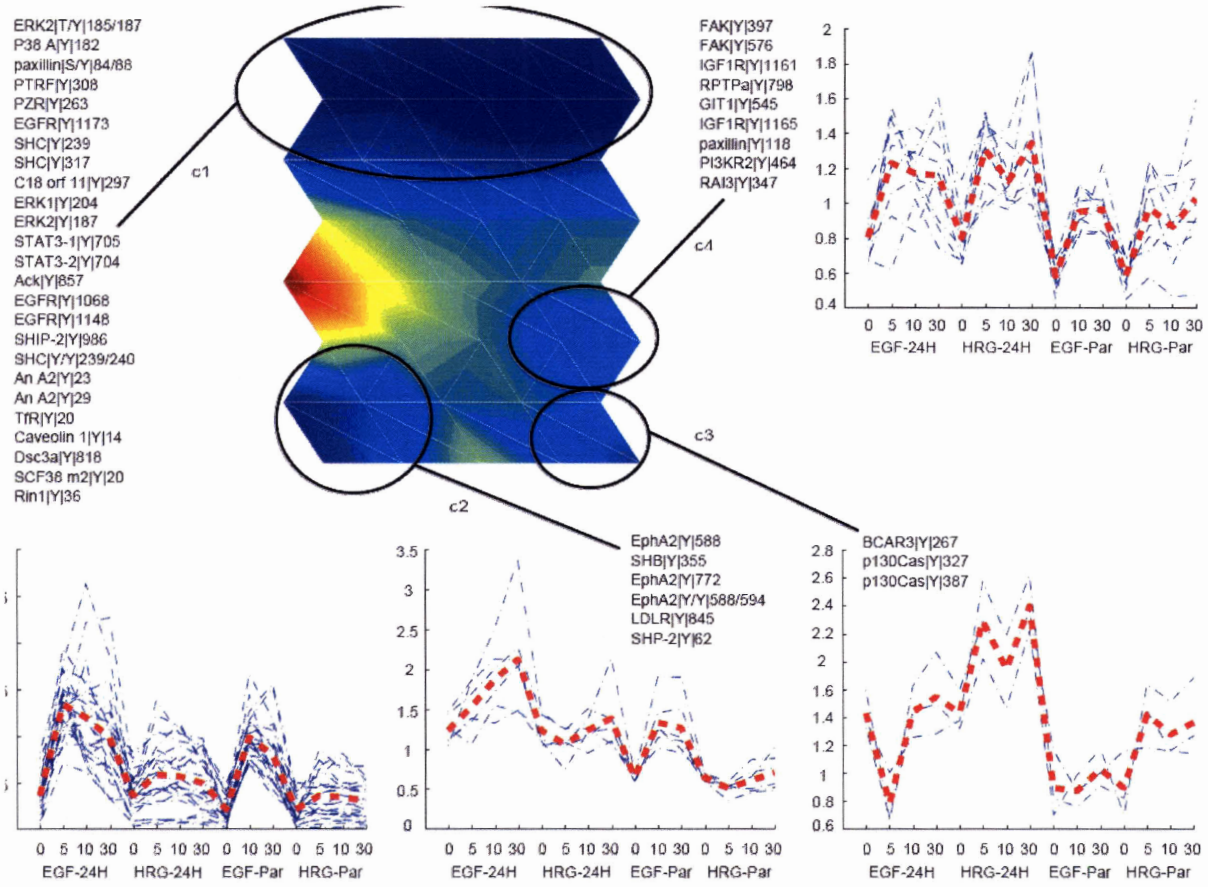


Figure 4-2: Self-Organizing Map to cluster phosphorylation sites across conditional and temporal space. Phosphorylation levels for 62 core phosphorylated peptides, which were detected and quantified at 4 time points for each of the 4 stimulation conditions and the 5 minute mix analysis, were normalized and placed into a self-organizing map (SOM). Four statistically significant clusters were identified in the SOM (circled bluish valleys surrounded by yellow "mountains"). The components for each of the four clusters are indicated in the figure along with the phosphorylation profiles (blue dashed lines) and an average profile for each cluster (red line) across conditions and time-points (16 dimensions).

By comparing the individual phosphorylation profiles (blue dashed lines) or average phosphorylation profile (red line) within each cluster, it is clear that the algorithm has clustered peptides whose phosphorylation level is increased under selected conditions, information which can be used to link phosphorylation sites within a cluster to activated receptor homo- or heterodimers. For instance, the first cluster (c1

in Figure 4-2) consists of 18 peptides whose phosphorylation levels, on average, are highest following EGF stimulation of P or 24H HMECs, conditions which would lead to activation of EGFR homodimers or EGFR-HER2 heterodimers, respectively. Correspondingly, most of the proteins in this cluster have been previously associated with proliferation and early response to EGF stimulation, including EGFR Y1068 and Y1148, STAT-3 Y705 (STAT-3 isoform 1) and Y704 (STAT-3 isoform 2), SHIP-2 Y986, SHC Y239, Y240 and Y317 and early effectors downstream of them, including MAP kinases ERK2 (phosphorylated at Y187 and at T185 and Y187), Erk1 Y204, and p38 □ Y182.

In comparison to c1, the three other clusters in the U-matrix (c2-c4) contain peptides whose phosphorylation appears to be regulated by HER2 activation, as phosphorylation levels on these peptides are highest when HER2 is overexpressed. More specifically, peptides in the second cluster (c2) are primarily phosphorylated downstream of EGFR-HER2 heterodimers, since the highest levels of phosphorylation occur in the 24H HMECs stimulated with EGF. This cluster consists of SHB Y355, SHP-2 Y62, LDL receptor Y845, and three EphA2 tyrosine sites, including the activation loop at Tyr 772 [29]. Of these proteins, only SHP-2 has been previously associated with EGFR activation [30]. Interestingly, each of the other proteins has been associated with the cell migration response to VEGF stimulation [31-33], a response which requires EGF autocrine signaling following VEGF stimulation in endothelial cells [34]. Our data links these phosphorylation sites to the EGFR family signaling network, and as demonstrated below, shows strong correlation between these phosphorylation sites and cell migration in response to EGF stimulation of 24H HMECs.

The third cluster (c3), contains 3 peptides that are primarily phosphorylated downstream of activated HER2-HER3 heterodimers, as implicated by maximal phosphorylation levels following HRG stimulation of 24H HMECs. This cluster includes phosphorylation of p130Cas at Y327 and Y387, and phosphorylation of BCAR3 at Y267. Both of these proteins are part of the same family and have been shown to regulate or be regulated by Src activity in the cell migration signaling network. Similar to c2 and c3, peptides in the fourth cluster (c4) are predominantly phosphorylated downstream of HER2, but are equally phosphorylated downstream of EGFR-HER2 or HER2-HER3 heterodimers or possibly activated through active HER2 homodimers (similar phosphorylation levels are seen in EGF or HRG stimulated 24H HMECs). This cluster includes phosphorylation of retinoic acid induced protein 3 (RAI3) Y347, paxillin Y118, GIT1 Y545, FAK Y397 and Y576, receptor protein tyrosine phosphatase alpha (RPTPa) Y798, PI3K Y464, and Insulin-like growth factor receptor (IGF-1R) Y1161 and Y1165. A complex containing GIT1 and paxillin has been shown to regulate cell motility, possibly through recruitment of PAK or PIX to the leading edge [35]; the role of tyrosine phosphorylation in this process has not been established. Other proteins in this cluster have also been shown to be co-regulated, as RPTPalpha has recently been shown to regulate integrin signaling through Src activation in a PI3K-dependent manner downstream of insulin activation [36], and Src activation is known to regulate tyrosine phosphorylation of FAK. It is worth noting that the phosphorylation sites on IGF-1R (Y1161 and Y1165) occur within a peptide that is homologous with the Insulin receptor. Although not conclusive, assignment to IGF-1R is based on the presence of a phosphorylation site on a peptide specific for IGF-1R in the larger data set and greater

levels of IGF-1R as compared to Insulin receptor on HMEC parental and 24H cells (H.S. Wiley, personal communication). Activation of IGR-1R has been found to be correlated with increased motility in a breast cancer cell line [37]; in our system, increased phosphorylation is most likely associated with autocrine release and binding of IGF-1 following stimulation of EGFR family members.

From these clusters, it is clear that analyzing phosphorylation sites across 16 dimensions of temporal and conditional space can reveal co-regulated phosphorylation sites, and that this information can be used to connect phosphorylation of specific sites to activation of EGFR family member dimers. Although clusters 3 and 4 both contain phosphorylation sites from proteins known to regulate migration, they are statistically distinct clusters whose differential response to exogenous stimuli may be due to protein sub-cellular localization, as proteins from cluster 4 are known to be associated with the membrane, while proteins in cluster 3 are primarily cytosolic. SOM-based data analysis can reveal interesting hypotheses and connectivity in the data, such as the potential role of GIT1 Y545 affecting cell migration or the inclusion of PTPRalpha, PI3K, and IGF-1R phosphorylation sites in a single cluster, but it is still necessary to relate phosphorylation and phenotypic data to solidify these hypotheses.

4.3.3 Cell proliferation and migration are differentially stimulated via EGFR and HER2

In order to correlate signaling data with cellular response, we quantified both cell migration and cell proliferation in the HMEC parental and 24H cell lines. Measurements

were obtained under serum-free, HRG (80 ng/ml), or EGF (100 ng/ml) stimulating conditions.

Cell migration was measured using a wound-closure assay adapted for a 96-well plate format and automated fluorescent imaging. Under all conditions investigated, the 24H cells moved more rapidly into the induced wound, thereby reducing the original wound area at a greater rate than the parental cells (Figure 4-3A-C). The wound closure trajectories were fit to a line, and the slopes were the input for partial least squares modeling. Interestingly, the greatest difference between the two cell lines occurs during EGF treatment. By comparison, HRG stimulation, while inducing a slightly higher rate of wound closure for 24H HMEC relative to P HMEC, does not seem to drive significantly enhanced migration relative to serum-free conditions.

Cell proliferation was measured by [^{3}H] thymidine uptake 25 hours after ligand stimulation. Figure 4-3D shows that EGF treatment increased thymidine uptake to a greater extent than did HRG treatment, but both treatments induced higher amounts of proliferation than seen in serum-free conditions. In direct contrast to the migration phenotype, there was no significant difference (i.e., $p >> 0.05$ in all cases) between the two cell types measured under any of the stimulation conditions. Thus, HER2 overexpression did not seem to facilitate enhanced proliferation in any of the conditions probed.

The fact that HER2 over-expression mediated differences in migration, particularly under EGF stimulating conditions, but did not do so for proliferation indicates that some of the signaling molecules differentially regulated by HER2 expression levels play a role in driving higher levels of migration while at the same time remaining

agnostic to cell proliferation. The linear regression modeling discussed below integrates our quantitative migration, proliferation, and signaling data to describe, among other things, a set of signaling molecules that is most relevant for the changes in migration induced by HER2 overexpression.

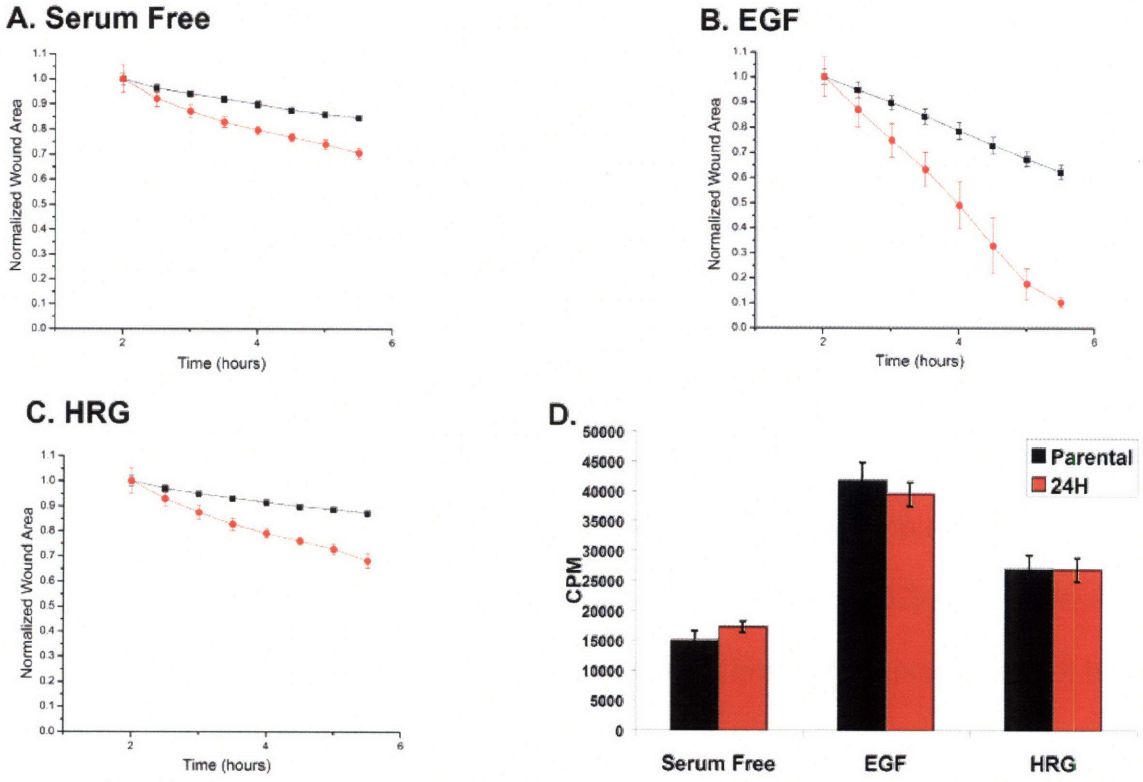


Figure 4-3: EGF and HRG drive migration and proliferation to varying extents in HMEC parental and HER2 over-expressing cells. Parental (black) and 24H (red) wound healing data shown for: (A) serum-free conditions, (B) treatment with 100 ng/ml EGF, and (C) treatment with 80 ng/ml HRG. (D) Proliferation, as measured by [3H] thymidine incorporation, for both cell types under the treatment conditions shown in (A)-(C).

4.3.4 Modulation of EGFR signaling by HER2

To assess the effect of increased HER2 expression levels in the canonical EGF activated pathway, the phosphorylation level for sites observed in EGF-stimulated 24H

cells was divided by the phosphorylation level for the same site and stimulation time in EGF-stimulated parental cells, producing a fold change in phosphorylation level for a given site and time. A subset of the proteins and phosphorylation sites within the canonical EGFR signaling network are shown in Figure 4-4A. As is evident from this figure, increased HER2 expression affects most phosphorylation sites on selected proteins in the EGFR signaling network, but not all phosphorylation sites on a given protein react equally to this perturbation. For instance, each of the multiple phosphorylation sites on EGFR exhibit different regulation at low or high HER2 expression levels, including increased phosphorylation of Y974 and decreased phosphorylation on Y1045 under HER2 overexpression as compared with basal HER2 expression. Both of these sites appear to regulate receptor internalization and degradation: Y974A or Y974F mutations have been shown to decrease receptor internalization rates [38], and Cbl (E3-ubiquitin ligase) binding to EGFR Y1045 is required for lysosomal sorting and receptor degradation [39]. Decreased phosphorylation at the Y1045 site should lead to decreased ubiquitination of activated EGFR, thereby providing a mechanism for observed increase in recycling of activated EGFR in the 24H cell line relative to parental 184A1 HMECs [40].

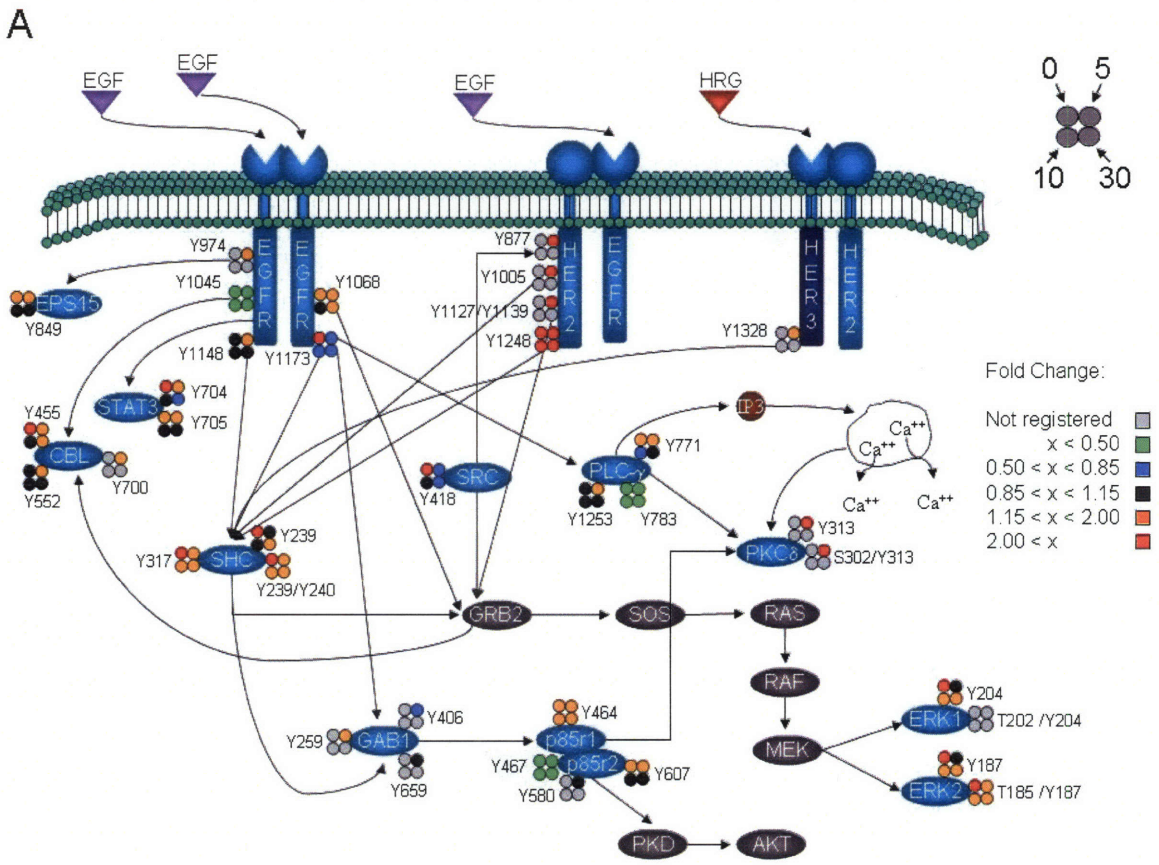
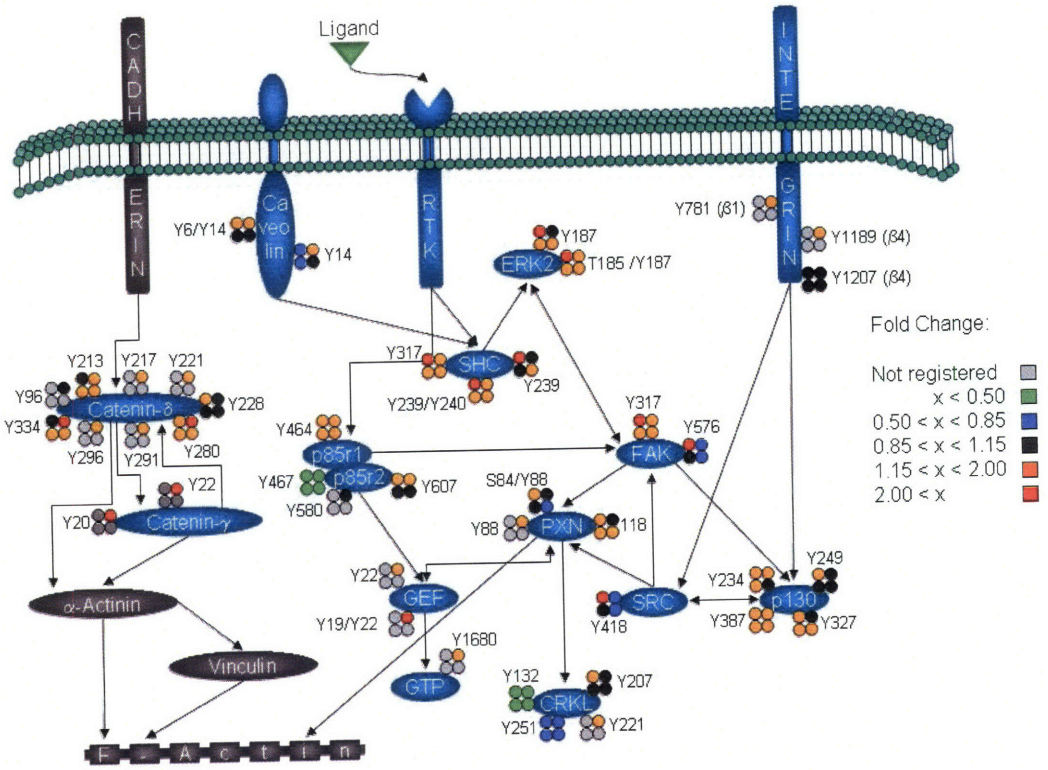


Figure 4-4 (A above and B below): Effect of increased HER2 expression on phosphorylation sites within the EGFR signaling network. Visualization of the fold change in phosphorylation level between 24H and parental cells stimulated with EGF provides a network view of the mechanistic effects of HER2 overexpression on (A) the canonical EGFR signaling cascades and (B) the cell migration associated signaling pathway.

B



It is also interesting to note the higher basal phosphorylation level for selected sites in the 24H cell lines relative to parental HMECs. Increased basal phosphorylation associated with HER2 overexpression could be mediated by HER2 autophosphorylation, cross-phosphorylation of EGFR in the absence of ligands, or even through increased autocrine stimulation of HER2 heterodimers. For most sites with high levels of basal phosphorylation in the 24H cells, stimulation with saturating concentrations of EGF typically resulted in a greater response in parental cells relative to 24H cells, such that after EGF stimulation for 5 minutes many of the sites had similar levels of phosphorylation in both cell lines. A specific example of this behavior is

provided by phosphorylation of the ERK2 activation loop (Y185 and T185/Y187). Much higher levels of basal phosphorylation were detected in the 24H cells relative to parental cells, but only a slight difference in phosphorylation at these sites remained after 5 minutes of EGF stimulation. These results are consistent with our proliferation data, in which no significant difference between 24H and parental cells under EGF stimulation was observed, in contrast to serum-free 24H cells, which had greater proliferation than parental HMECs ($p < 0.1$). These results are also consistent with previous literature demonstrating ERK2's role as a potent activator of proliferation [41]. In fact, many protein phosphorylation sites associated with proliferation behave similarly to ERK2 in our system, including STAT-3 Y705 (and Y704 from STAT-3 isoform 2), EGFR Y1173, and SHC Y239, Y240, and Y317.

Since HER2 over-expression has been found to affect cell motility in our study here as well as in previous work [42], the effect on protein phosphorylation in a subset of the pathways related to cell migration is shown in Figure 4-4B. From this figure, it is clear that there is an overall increase in phosphorylation of many of these pathway proteins in the presence of HER2 overexpression. Perhaps most striking is the increase in phosphorylation for all of the phosphorylation sites detected on catenin-δ and catenin-γ. Catenins are known to interact with E-Cadherin, the main cell-cell adhesion protein in epithelial cells [43]. Catenin-δ, a member of the p120 catenin family, has been shown to regulate E-Cadherin turnover by modulating its internalization and degradation rate [43], stabilizing it on the cell surface when bound to it. HER2 overexpression in breast carcinomas inhibits the transcription of E-Cadherin [44] and has also been found to destabilize the catenin-cadherin complex, leading to decreased adhesion [45]. Also,

EGFR has been found to directly phosphorylate p120 catenin at Y228 [46], and EGFR inhibition was found to promote assembly of cell adhesions [47]. The accumulated evidence of these reports is consistent with our data as it argues that Src- or RTK-related phosphorylation of catenin-δ leads to separation of the catenin-cadherin complex, resulting in destabilization of E-Cadherin and an increase in cell migration. Tyrosine phosphorylation of catenin-γ has also been related to loss of cell-cell adhesions in EGF-stimulated, E-Cadherin positive, cervical cancer cells [48]. Our data shows high levels of phosphorylation of catenin-δ and γ in HER2 over-expressing cells under EGF stimulation, directly contributing to loss of cell adhesion and an increase in cell motility.

Although catenin phosphorylation and loss of E-cadherin-based cell-cell adhesion may provide part of the migratory response following EGF stimulation in 24H cells, increased phosphorylation of many additional components of the cell migration pathways are also likely to be contributing to increased migration of these cells relative to parental cells. For instance, FAK, Src, Paxillin and p130Cas have all been shown to interact with each other, to localize at the focal adhesions, and to play a fundamental role in the actin cytoskeleton reorganization and motility pathways [49]. Phosphorylation of paxillin (Y31) and (Y118) is known to regulate membrane spreading and ruffling in cell migration and adhesion [50]. FAK and Src are two of the major kinases involved in cell migration. FAK (Y397) is the major autophosphorylation site and acts as a docking site for Src and PI3K [51]; this site is also necessary for both p130cas and paxillin phosphorylation in response to FAK expression [51]. Src Y418 is a major Src autophosphorylation site, whose phosphorylation results in self-activation

[52]. Src can mediate FAK related migration by further phosphorylation of paxillin or p130Cas, which may result in focal adhesion turnover and/or lamellipodia and filipodia formation via either the RAC, JNK pathway or the RhoGAP pathway [53]. We conclude that the effect of increased HER-2 expression levels is partly increased interaction of FAK, Src, P130Cas and paxillin, a small but noticeable increase in ERK2 activity and the upregulation of catenin phosphorylation and thus E-cadherin turnover, all combining to drive an increased migratory response in the 24H cell line (as in Figure 4-3).

4.3.5 HRG vs EGF stimulation in the presence of HER2

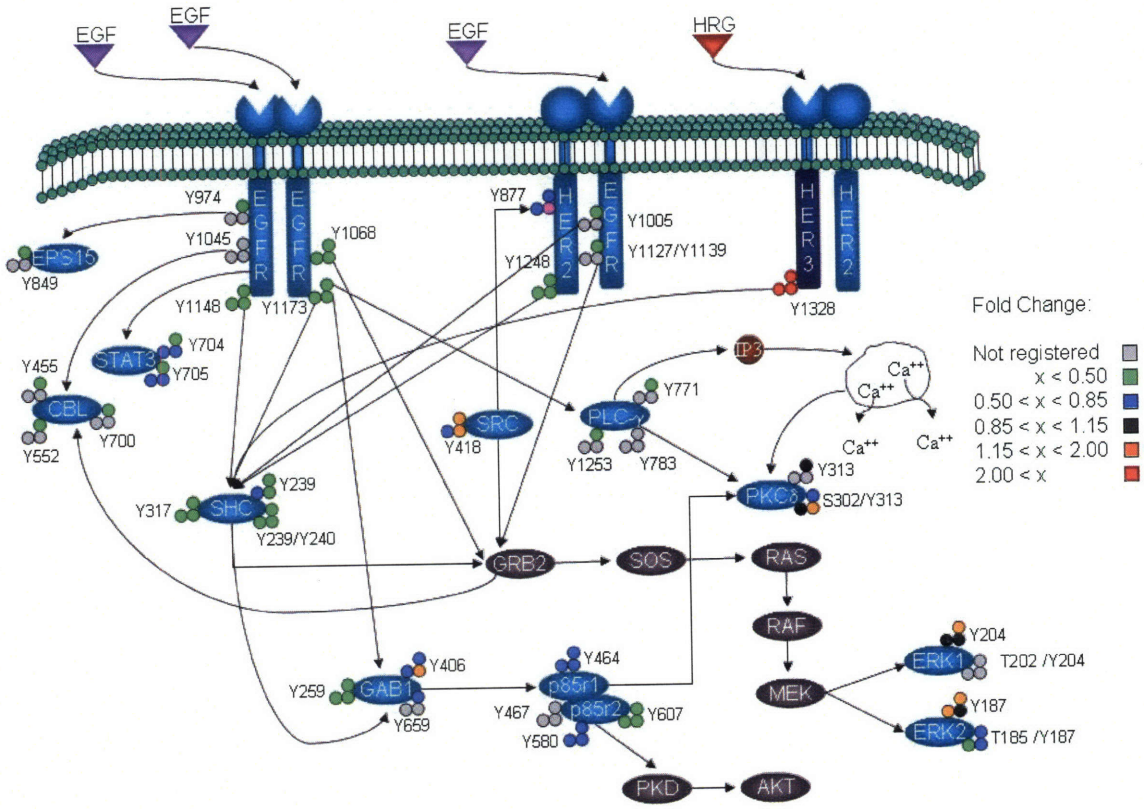
In addition to investigating the effects of HER2 on these pathways in Figure 4-5, our data also enabled comparison of signaling downstream of activated HER2:HER3 vs. EGFR:HER2 heterodimers, resulting from stimulation of the 24H cells with HRG and EGF, respectively. For this analysis, the phosphorylation level resulting from HRG stimulation of 24H cells was divided by the phosphorylation level resulting from EGF stimulation of 24H cells to produce the fold change ratio for each phosphorylation site in the canonical signaling pathways (Figure 4-5A). As expected, stimulation with HRG instead of EGF caused a significant increase in HER3 phosphorylation and a large decrease in EGFR phosphorylation. Perhaps not surprisingly given the relative receptor expression levels (20,000-30,000 copies/cell for HER3 vs. 200,000 copies/cell for EGFR) and the kinase-dead nature of HER3, stimulation with saturating concentrations of HRG resulted in decreased phosphorylation of almost all downstream proteins as compared to stimulation with saturating concentrations of EGF (Figure 4-5A).

Specifically, most of the downstream effectors which lead to proliferation (STAT-3 Y705 (Y704), ERK T202/Y204, all three phosphorylation sites on SHC) were phosphorylated to a lesser degree under HRG stimulation, correlating with decreased proliferation in HRG stimulated 24H cells as compared to EGF stimulated 24H cells (as in Figure 4-3). Other sites primarily associated with migration (Src, PKC-□, and PI3K) do not show the same decrease in phosphorylation when comparing HRG to EGF stimulation of these cells. In fact, both Src and PKC-□ show an increase in phosphorylation under HRG activation at 5 and 10 minutes respectively. Together with our response data, these results indicate that signaling distinct from Src, PKC-□, and PI3K controlled pathways may govern cell migration in our system.

To analyze the effect of HRG stimulation on the cell migration signaling network, signaling downstream of activated HER2:HER3 vs. EGFR:HER2 heterodimers was also compared for selected proteins within a subset of the pathways related to cell motility (Figure 4-5B). Comparing phosphorylation levels for specific sites under different stimulation conditions provides a potential hypothesis to explain the mechanism underlying the quantitative phenotypic migration measurements, in which EGF stimulation resulted in a dramatic increase in migration relative to HRG stimulation of 24H cells (see Figure 4-3). Given the difference in migration rates between these two cell states, it is not surprising to note that protein phosphorylation sites on almost all effectors of migration have decreased phosphorylation levels following HRG stimulation as compared to EGF stimulation. However, in contrast to most proteins in the motility network, tyrosine phosphorylation levels on FAK, p130Cas, Src and paxillin were increased following HRG stimulation compared to EGF stimulation, indicating that the

slight migratory effect of HRG stimulation of 24H cells is directed primarily through amplified phosphorylation of a very specific pathway driven through these four proteins.

A



B

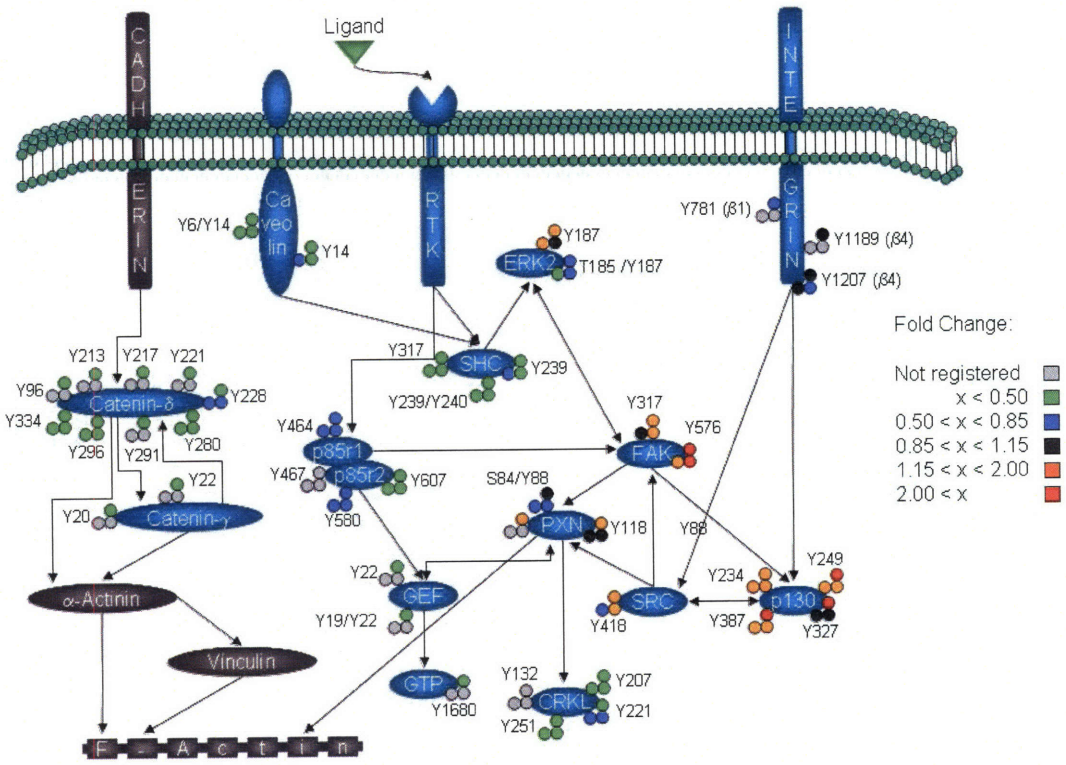


Figure 4-5: Effect of HRG stimulation versus EGF stimulation in high HER2 expressing cells. Visualization of the fold change in phosphorylation level between HRG-stimulated and EGF-stimulated 24H cells provides a network view of the mechanistic effects underlying differential response to distinct growth factor stimuli on (A) the canonical EGFR signaling cascades and (B) the cell migration associated signaling pathway.

Taken together, the data from Figures 4-4 and 4-5 demonstrate that 24H cell migration can be induced by either broad upregulation of the migratory pathway, including loss of cell adhesions (24H cells under EGF stimulation) or, to a much lesser extent, by more intense upregulation of a restricted subset of the migratory pathway (as in 24H cells stimulated with HRG). By comparison, 24H cell proliferation was induced to a greater extent by EGF stimulation as determined by phenotypic assay;

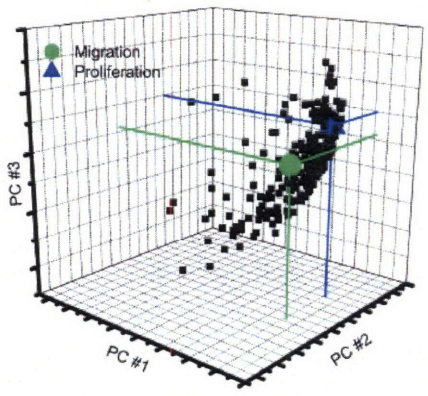
correspondingly, phosphorylation levels were higher on almost all proliferation-associated proteins on 24H cells stimulated with EGF as compared to 24H cells stimulated with HRG. Also consistent with the phenotypic data, the effect of increasing HER2 expression levels was most dramatic for proteins in the cell migration signaling network, as most proteins displayed a sustained increase in phosphorylation levels, but was much less dramatic for proteins in the proliferation-associated network, as increased basal phosphorylation levels in 24H cells were not sustained following EGF stimulation.

4.3.6 Linear modeling correlates signals with cell function

We have constructed a model using partial least-squares regression (PLSR), a method we have previously found to be effective in relating cell signaling data to cell behavioral response data in a quantitative and integrative manner [54]. Information obtained through our proteomic studies was represented in an *M x N* matrix (the X-block), where M is the number of conditions investigated, and N is the number of peptide metrics measured. An entry in the matrix with coordinates (*i,j*) represents the column *j* metric (i.e. ERK Y187 phosphorylation at 5 minutes) measured under the row *i* condition (i.e. parental cell line treated with EGF). For each condition, the metrics included in the model were phosphorylation measurements at 5, 10, and 30 minutes in addition to the integral of this time course (with integrals being used as a measurement for the 'net' phosphorylation over the 30 minute time course). Cell behavior measurements comprised an *M x P* matrix (the Y-block), where M is again the number

of conditions and P is the number of behavior measurements obtained. Partial least-squares regression analysis produced a vector of coefficients indicating the importance of each signaling metric with respect to cellular behavior. In addition, PLSR provided a reduced-dimension map, with axes defined as linear combinations of our original signaling metrics (Figure 4-6A), on which both signals and cellular behavior can be represented. Figure 4-6A shows that our original dataset, consisting of 248 dimensions (i.e., 248 protein signal metrics), has been reduced to 3 dimensions using PLSR, each of which incorporates a quantitative combination of multiple signals. The projection of an individual signal in the direction of a given cellular behavior in the PLS space determines how important the phosphorylation signal is to the behavior. In Figure 4-6B, we list the top 20 signals that positively correlate most strongly with each cell behavior. Importantly, even though we can identify small sets of variables that correlate strongly with each cellular output, 148 out of the 248 protein metrics had a variable importance for projection (VIP) value of greater than 1, indicating that these 148 protein metrics play an important role in our model (see Methods for VIP calculation). This highlights the great advantage of proteome measurements that quantitatively capture dynamic information flow through a large number of nodes. Our model was validated through cross-validation and had a goodness of prediction (Q^2) of 0.89 (see Methods).

A.



B.

Migration		Proliferation	
Phosphorylation Site	Measurement	Phosphorylation Site	Measurement
SHP-2 Y 62	30 minutes	Dsc3a Y 818	10 minutes
An A2 Y 237	10 minutes	Dsc3a Y 818	30 minutes
An A2 Y 237	30 minutes	EGFR Y 1173	10 minutes
HER2 Y 1248	5 minutes	EGFR Y 1173	integral
Erbin Y 1104	5 minutes	EGFR Y 1173	30 minutes
HER2 Y 1248	10 minutes	Dsc3a Y 818	integral
HER2 Y 1248	30 minutes	IGF1R Y 1165	30 minutes
GRF1 Y 1105	5 minutes	EGFR Y 1173	5 minutes
LDLR Y 845	30 minutes	paxillin S/Y 84/88	30 minutes
HER2 Y 1248	integral	CrkL Y 132	10 minutes
SHP-2 Y 62	10 minutes	GIT1 Y 545	5 minutes
An A2 Y 237	5 minutes	paxillin S/Y 84/88	5 minutes
SHB Y 355	5 minutes	Catenin d1 Y 228	30 minutes
LDLR Y 845	5 minutes	paxillin S/Y 84/88	integral
EphA2 Y/Y 588/594	5 minutes	Src Y 418	integral
LDLR Y 845	10 minutes	Catenin d1 Y 228	integral
EphA2 Y 575	30 minutes	EGFR Y 1148	10 minutes
KIAA1217 Y 393	5 minutes	IGF1R Y 1161	10 minutes
An A2 Y 23	5 minutes	CrkL Y 132	integral
SHP-2 Y 62	5 minutes	paxillin S/Y 84/88	10 minutes

Figure 4-6: Partial least squares regression correlates 248 protein metrics to cell migration and proliferation. (A) Visual representation of a reduced (3) dimension graph showing all 248 protein measurements and the cellular outputs with cell migration (green), cell proliferation (blue), and protein signals (black). (B) List of the top 20 signal-behavior covariates for both migration and proliferation.

From the top 20 signal behavior co-variates for proliferation and migration in our model (Figure 4-6B), HER Y1248 appears to be the main upstream activator of migration while proliferation seems to be activated by EGFR Y1173 and to a lesser extent EGFR Y1148. This finding correlates with literature showing that the presence of HER2 Y1248 is necessary to induce migration of breast cancer cell lines [55] whereas both EGFR Y1148 and Y1173 are Shc binding sites and thus are able to activate the ERK signaling pathway, thereby driving proliferation [56]. In this model, SHP-2 Y62 and Annexin A2 Y237, among others, also appear to be highly correlated with migration. Both of these proteins have previously been implicated in regulating migration, as SHP-2 has been shown to promote migration in MCF-7 cells by inducing loss of E-Cadherin expression and production of matrix metalloproteinase 9[57], while Annexin A2 has

been shown to mediate mitogenesis, cell migration and loss of focal adhesions when activated by tenascin-c [58, 59] and it has also been identified as a major substrate for EGFR phosphorylation [60]. Interestingly, treating cells with withaferin, a novel inhibitor of Annexin A2, has recently been shown to decrease migration, further validating the correlation between Annexin A2 and cell migration [61].

In addition to EGFR Y1173 and Y1148, desmocollin (Dsc3a) Y818, IGF-1R Y1165, and catenin-δ Y228 (among others) are most strongly correlated with proliferation. IGF-1R is known to induce cell proliferation and survival through activation of the MAPK and PI3K pathways [62], and has also been associated with EGFR signaling in Tamoxifen resistant breast cancer cell lines, potentiating their mitogenic strength [10]. The correlation of Dsc3a and catenin-δ to cellular proliferation is initially surprising, as both of these proteins are associated with cell adhesions, leading to the expectation that they should correlate to migration instead of proliferation. However, EGFR inhibition has been shown to increase the presence of desmocollin and desmoglein at cell-cell borders [47], and EGFR has also been shown to signal to E-cadherin complexes through phosphorylation of catenin-δ Y228 [46]. Since the conditional and temporal profiles for both of these sites (Dsc3a Y818 and catenin-δ Y228) display strong similarity to that observed for EGFR Y1173, it can be argued that they are either directly or closely downstream of EGFR and that phosphorylation of these sites may destabilize cell-cell adhesions, a step needed for both proliferation and migration. Further biological experiments are clearly needed to more firmly establish the functional roles of these proteins in regulating cellular proliferation under these stimulation conditions.

It is worth noting that a large proportion of the phosphorylation events contained in our regression model influence both migration and proliferation to some extent. Importantly, the model provides a quantitative metric describing the strength of correlation for each phosphorylation site relative to migration or proliferation; data which can be used to design further experiments to specifically perturb selected biological outcomes -- i.e., to inhibit migration one might want to target HER2, SHP-2 and EphA2. This type of data-driven modeling can also provide insight to the functionality of unknown proteins such as KIAA127; in our model, phosphorylation of this protein at Y393 correlates strongly to migration. Interestingly, KIAA1217 is a novel protein highly homologous to p140CAP (p130Cas-associated protein), which has been shown to be tyrosine phosphorylated in response to EGF stimulation and to participate in actin cytoskeleton organization and cell spreading [63].

4.4 Conclusions

Combining large-scale quantitative analysis of tyrosine phosphorylation with quantitative phenotypic measurements has provided the means with which to understand the relationship between these phosphorylation events and their relation to cellular responses. The findings we have discovered here illustrate how HER2 over-expression influences signaling network activities important for governing cell proliferation and migration behavior, common to and distinct between EGFR-binding ligand and HER3-binding ligand conditions. This quantitative information offers unusual opportunity for understanding prospective drug effects in a network-wide manner, and

may offer novel targets for intervening in biological processes downstream of activated receptor tyrosine kinases.

References

1. Mendelsohn, J. and J. Baselga, *Status of epidermal growth factor receptor antagonists in the biology and treatment of cancer.* J Clin Oncol, 2003. **21**(14): p. 2787-99.
2. Yarden, Y. and M.X. Sliwkowski, *Untangling the ErbB signalling network.* Nat Rev Mol Cell Biol, 2001. **2**(2): p. 127-37.
3. Hynes, N.E. and H.A. Lane, *ERBB receptors and cancer: the complexity of targeted inhibitors.* Nat Rev Cancer, 2005. **5**(5): p. 341-54.
4. Lemmon, M.A., *The EGF receptor family as therapeutic targets in breast cancer.* Breast Dis, 2003. **18**: p. 33-43.
5. Carlsson, J., et al., *HER2 expression in breast cancer primary tumours and corresponding metastases. Original data and literature review.* Br J Cancer, 2004. **90**(12): p. 2344-8.
6. Citri, A., K.B. Skaria, and Y. Yarden, *The deaf and the dumb: the biology of ErbB-2 and ErbB-3.* Exp Cell Res, 2003. **284**(1): p. 54-65.
7. Ross, J.S., et al., *The Her-2/neu gene and protein in breast cancer 2003: biomarker and target of therapy.* Oncologist, 2003. **8**(4): p. 307-25.
8. Harari, D. and Y. Yarden, *Molecular mechanisms underlying ErbB2/HER2 action in breast cancer.* Oncogene, 2000. **19**(53): p. 6102-14.

9. Holbro, T., et al., *The ErbB2/ErbB3 heterodimer functions as an oncogenic unit: ErbB2 requires ErbB3 to drive breast tumor cell proliferation.* Proc Natl Acad Sci U S A, 2003. **100**(15): p. 8933-8.

10. Knowlden, J.M., et al., *Elevated levels of epidermal growth factor receptor/c-erbB2 heterodimers mediate an autocrine growth regulatory pathway in tamoxifen-resistant MCF-7 cells.* Endocrinology, 2003. **144**(3): p. 1032-44.

11. Pinkas-Kramarski, R., et al., *Diversification of Neu differentiation factor and epidermal growth factor signaling by combinatorial receptor interactions.* Embo J, 1996. **15**(10): p. 2452-67.

12. Aguilar, Z., et al., *Biologic effects of heregulin/neu differentiation factor on normal and malignant human breast and ovarian epithelial cells.* Oncogene, 1999. **18**(44): p. 6050-62.

13. Atlas, E., et al., *Heregulin is sufficient for the promotion of tumorigenicity and metastasis of breast cancer cells in vivo.* Mol Cancer Res, 2003. **1**(3): p. 165-75.

14. Tsai, M.S., et al., *Blockage of heregulin expression inhibits tumorigenicity and metastasis of breast cancer.* Oncogene, 2003. **22**(5): p. 761-8.

15. Yao, J., et al., *Multiple signaling pathways involved in activation of matrix metalloproteinase-9 (MMP-9) by heregulin-beta1 in human breast cancer cells.* Oncogene, 2001. **20**(56): p. 8066-74.

16. Zhang, Y., et al., *Time-resolved Mass Spectrometry of Tyrosine Phosphorylation Sites in the Epidermal Growth Factor Receptor Signaling Network Reveals Dynamic Modules.* Mol Cell Proteomics, 2005. **4**(9): p. 1240-1250.

17. Stampfer, M.R. and J.C. Bartley, *Induction of transformation and continuous cell lines from normal human mammary epithelial cells after exposure to benzo[a]pyrene.* Proc Natl Acad Sci U S A, 1985. **82**(8): p. 2394-8.

18. Hendriks, B.S., et al., *Coregulation of epidermal growth factor receptor/human epidermal growth factor receptor 2 (HER2) levels and locations: quantitative analysis of HER2 overexpression effects.* Cancer Res, 2003. **63**(5): p. 1130-7.

19. Kohonen, T., *Self-Organizing Maps.* 3 ed. Information Sciences. 2001, Berlin, Germany: Springer. 521.

20. Vesanto, J., et al., *SOM toolbox for Matlab 5*, in *Technical Report A57*. 2000, Helsinki University of Technology, Finland.

21. Ultsch, A. and H. Siemon, *Exploratory data analysis: Using Kohonen networks on transputers.* 1989.

22. Hautaniemi, S., et al., *Analysis and Visualization of Gene Expression Microarray Data in Human Cancer Using Self-Organizing Maps.* Machine Learning, 2003. **52**(1-2): p. 45-66.

23. Gaudet, S., et al., *A Compendium of Signals and Responses Triggered by Prodeath and Prosurvival Cytokines.* Mol Cell Proteomics, 2005. **4**(10): p. 1569-1590.

24. Janes, K.A., et al., *Cue-signal-response analysis of TNF-induced apoptosis by partial least squares regression of dynamic multivariate data.* J Comput Biol, 2004. **11**(4): p. 544-61.

25. Eriksson, L., et al., *Multi- and Megavariate Data Analysis Principles and Applications.* 2001.

26. Hendriks, B.S., et al., *Parsing ERK activation reveals quantitatively equivalent contributions from epidermal growth factor receptor and HER2 in human mammary epithelial cells.* J Biol Chem, 2005. **280**(7): p. 6157-69.

27. Schulze, W.X., L. Deng, and M. Mann, *Phosphotyrosine interactome of the ErbB-receptor kinase family.* Mol Syst Biol, 2005. **1**(1): p. msb4100012-E1.

28. Xia, L., et al., *Identification of both positive and negative domains within the epidermal growth factor receptor COOH-terminal region for signal transducer and activator of transcription (STAT) activation.* J Biol Chem, 2002. **277**(34): p. 30716-23.

29. Kinch, M.S. and K. Carles-Kinch, *Overexpression and functional alterations of the EphA2 tyrosine kinase in cancer.* Clin Exp Metastasis, 2003. **20**(1): p. 59-68.

30. Qu, C.K., et al., *Genetic evidence that Shp-2 tyrosine phosphatase is a signal enhancer of the epidermal growth factor receptor in mammals.* Proc Natl Acad Sci U S A, 1999. **96**(15): p. 8528-33.

31. Cheng, N., et al., *Blockade of EphA receptor tyrosine kinase activation inhibits vascular endothelial cell growth factor-induced angiogenesis.* Mol Cancer Res, 2002. **1**(1): p. 2-11.

32. Holmqvist, K., et al., *The Shb adaptor protein causes Src-dependent cell spreading and activation of focal adhesion kinase in murine brain endothelial cells.* Cell Signal, 2003. **15**(2): p. 171-9.

33. Prager, G.W., et al., *Vascular endothelial growth factor receptor-2-induced initial endothelial cell migration depends on the presence of the urokinase receptor.* Circ Res, 2004: p. Epub 2004 May 6.

34. Semino, C.E., R.D. Kamm, and D.A. Lauffenburger, *Autocrine EGF receptor activation mediates endothelial cell migration and vascular morphogenesis induced by VEGF under interstitial flow.* Exp Cell Res, 2006: p. Epub 2005 Dec 7.

35. Manabe, R., et al., *GIT1 functions in a motile, multi-molecular signaling complex that regulates protrusive activity and cell migration.* J Cell Sci, 2002. **115**(Pt 7): p. 1497-510.

36. Vulin, A.I., K.K. Jacob, and F.M. Stanley, *Integrin activates receptor-like protein tyrosine phosphatase alpha, Src, and Rho to increase prolactin gene expression through a final phosphatidylinositol 3-kinase/cytoskeletal pathway that is additive with insulin.* Endocrinology, 2005. **146**(8): p. 3535-46.

37. Zhang, X., et al., *Multiple Signaling Pathways are Activated During Insulin-like Growth Factor-I (IGF-I) Stimulated Breast Cancer Cell Migration.* Breast Cancer Res Treat, 2005. **93**(2): p. 159-68.

38. Sorkin, A., et al., *Epidermal growth factor receptor interaction with clathrin adaptors is mediated by the Tyr974-containing internalization motif.* J Biol Chem, 1996. **271**(23): p. 13377-84.

39. Grovdal, L.M., et al., *Direct interaction of Cbl with pTyr 1045 of the EGF receptor (EGFR) is required to sort the EGFR to lysosomes for degradation.* Exp Cell Res, 2004. **300**(2): p. 388-95.

40. Hendriks, B.S., et al., *Quantitative analysis of HER2-mediated effects on HER2 and epidermal growth factor receptor endocytosis: distribution of homo- and*

heterodimers depends on relative HER2 levels. J Biol Chem, 2003. **278**(26): p. 23343-51.

41. Cobb, M.H., et al., *The mitogen-activated protein kinases, ERK1 and ERK2.* Semin Cancer Biol, 1994. **5**(4): p. 261-8.

42. Spencer, K.S., et al., *ErbB2 is necessary for induction of carcinoma cell invasion by ErbB family receptor tyrosine kinases.* J Cell Biol, 2000. **148**(2): p. 385-97.

43. Davis, M.A., R.C. Ireton, and A.B. Reynolds, *A core function for p120-catenin in cadherin turnover.* J Cell Biol, 2003. **163**(3): p. 525-34.

44. D'Souza, B. and J. Taylor-Papadimitriou, *Overexpression of ERBB2 in human mammary epithelial cells signals inhibition of transcription of the E-cadherin gene.* Proc Natl Acad Sci U S A, 1994. **91**(15): p. 7202-6.

45. Jawhari, A.U., M.J. Farthing, and M. Pignatelli, *The E-cadherin/epidermal growth factor receptor interaction: a hypothesis of reciprocal and reversible control of intercellular adhesion and cell proliferation.* J Pathol, 1999. **187**(2): p. 155-7.

46. Mariner, D.J., M.A. Davis, and A.B. Reynolds, *EGFR signaling to p120-catenin through phosphorylation at Y228.* J Cell Sci, 2004. **117**(Pt 8): p. 1339-50.

47. Lorch, J.H., et al., *Epidermal growth factor receptor inhibition promotes desmosome assembly and strengthens intercellular adhesion in squamous cell carcinoma cells.* J Biol Chem, 2004. **279**(35): p. 37191-200.

48. Moon, H.S., et al., *Expression and tyrosine phosphorylation of E-cadherin, beta- and gamma-catenin, and epidermal growth factor receptor in cervical cancer cells.* Gynecol Oncol, 2001. **81**(3): p. 355-9.

49. 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.

50. Tsubouchi, A., et al., *Localized suppression of RhoA activity by Tyr31/118-phosphorylated paxillin in cell adhesion and migration.* J Cell Biol, 2002. **159**(4): p. 673-83.

51. Cary, L.A. and J.L. Guan, *Focal adhesion kinase in integrin-mediated signaling.* Front Biosci, 1999. **4**: p. D102-13.

52. Roskoski, R., Jr., *Src protein-tyrosine kinase structure and regulation.* Biochem Biophys Res Commun, 2004. **324**(4): p. 1155-64.

53. Playford, M.P. and M.D. Schaller, *The interplay between Src and integrins in normal and tumor biology.* Oncogene, 2004. **23**(48): p. 7928-46.

54. 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.

55. Dittmar, T., et al., *Induction of cancer cell migration by epidermal growth factor is initiated by specific phosphorylation of tyrosine 1248 of c-erbB-2 receptor via EGFR.* Faseb J, 2002. **16**(13): p. 1823-5.

56. Okabayashi, Y., et al., *Tyrosines 1148 and 1173 of activated human epidermal growth factor receptors are binding sites of Shc in intact cells.* J Biol Chem, 1994. **269**(28): p. 18674-8.

57. Wang, F.M., et al., *SHP-2 promoting migration and metastasis of MCF-7 with loss of E-cadherin, dephosphorylation of FAK and secretion of MMP-9 induced by IL-1beta in vivo and in vitro.* Breast Cancer Res Treat, 2005. **89**(1): p. 5-14.

58. Liu, J.W., et al., *Annexin II expression is reduced or lost in prostate cancer cells and its re-expression inhibits prostate cancer cell migration.* Oncogene, 2003. **22**(10): p. 1475-85.

59. Chung, C.Y., J.E. Murphy-Ullrich, and H.P. Erickson, *Mitogenesis, cell migration, and loss of focal adhesions induced by tenascin-C interacting with its cell surface receptor, annexin II.* Mol Biol Cell, 1996. **7**(6): p. 883-92.

60. Rothhut, B., *Participation of annexins in protein phosphorylation.* Cell Mol Life Sci, 1997. **53**(6): p. 522-6.

61. Falsey, R.R., et al., *Actin microfilament aggregation induced by withaferin A is mediated by annexin II.* Nat Chem Biol, 2006: p. Epub 2005 Dec 11.

62. Wang, Y., et al., *Inhibition of insulin-like growth factor-I receptor (IGF-IR) signaling and tumor cell growth by a fully human neutralizing anti-IGF-IR antibody.* Mol Cancer Ther, 2005. **4**(8): p. 1214-21.

63. Di Stefano, P., et al., *P130Cas-associated protein (p140Cap) as a new tyrosine-phosphorylated protein involved in cell spreading.* Mol Biol Cell, 2004. **15**(2): p. 787-800.

Chapter 5 Modeling HER2 effects on cell behavior from mass spectrometry phosphotyrosine data

This chapter builds from the dataset acquired in Chapter 4 to demonstrate the use of PLSR-based models to systematically characterize the signaling events that regulate cell migration and proliferation when HER2 is overexpressed under a variety of ligand treatment conditions.

5.1 Introduction

Recent advances in mass spectrometry have enabled the extensive characterization of intracellular signaling networks [1, 2]. Coupled with the increasing appreciation that cell behavior is governed by a network of signaling events, these advances have been used to identify novel elements of network activation giving rise to cell behavior. Identification of such elements in the past has largely been accomplished in a non-structured way through the manual parallel comparison of fold-change phosphorylation and cell phenotype [3, 4]. We sought to use a mathematical formalism based on linear mapping to draw predictive connections between cell behavior (migration and proliferation) and a mass spectrometry dataset describing changes in intracellular tyrosine phosphorylation as human epidermal growth factor receptor 2 (HER2) was overexpressed under a variety of ligand treatment conditions.

HER2, a member of the ErbB family of receptors, is overexpressed in ~30% of breast cancer patients and correlates with poor prognosis and high invasiveness [5]. Other members of the ErbB receptor family include epidermal growth factor receptor

(EGFR), human epidermal growth factor receptor 3 (HER3), and human epidermal growth factor receptor 4 (HER4). These receptors give rise to one of the most extensively studied signaling networks in biology through a variety of ligand binding and dimerization schemes [6, 7]. Epidermal growth factor (EGF) and heregulin (HRG), two ErbB family ligands, have been shown to induce both proliferation and migration to varying extents in breast cancer cells, although the signaling mechanisms responsible for this are not fully understood [8, 9]. EGF predominantly binds EGFR to induce both EGFR homodimers and EGFR-HER2 heterodimers, whereas heregulin (HRG) predominantly binds HER3 and HER4, inducing HER2-HER3 and HER2-HER4 heterodimers. To obtain a dynamic and quantitative description of intracellular signaling in response to treatment with EGF or HRG and changing HER2 levels, we employed a mass spectrometry approach that measured levels of tyrosine phosphorylation. Cell migration and proliferation were also quantified under the same treatment conditions [10]. Partial least squares regression (PLSR), a technique previously shown to be useful for the creation of signal-response models based on highly dimensional datasets, was used to correlate phosphorylation events to both migration and proliferation [10].

In this study, we demonstrate the use of PLSR-based models to systematically characterize the signaling events that regulate cell migration and proliferation when HER2 is overexpressed under a variety of ligand treatment conditions. Specifically, we derive lists of the most important phenotypically-relevant proteins characterizing each of 30 possible transitions between our six cell conditions (EGF, HRG, and serum-free treatments in both low and high HER2-expressing cells). Inspection of the lists reveals both regulatory signaling cascades consistent with known HER2 biology and novel

hypotheses. Using a conceptually similar procedure, we also derived lists of proteins that uniquely correlated with either migration or proliferation, postulating that these proteins serve as migration- or proliferation- specific signals in our system. Finally, we analyzed the PLSR model to derive a subset of phosphorylation sites most informative for the quantitative prediction of migration and proliferation. We identified nine phosphosites (signals) on six proteins from the original 62 phosphosites (signals), and showed that a model based on only those nine sites had a goodness of fit to experimental data similar to the full model. We identify the nine signals as a 'network gauge,' a subset of molecules in the vast network of signaling molecules that together serve as a sensitive readout for cellular response. The non-obvious nature of the nine selected signals highlights the complexity of the network and the usefulness of the modeling approach. Analysis of the network gauge suggests that two elements of network architecture, endocytosis and phosphoinositide 3-kinase (PI3K)-related signaling, are highly informative loci for the control of proliferation and migration. Importantly, models constructed from both the full and network gauge signaling data that were trained only on data from a low HER2-expressing cell line predicted levels of migration and proliferation in a HER2-overexpressing cell line for both EGF and HRG treatments. This suggests that both cell types process information in the signaling network according to the same set of multi-linear rules.

5.2 Materials and Methods

5.2.1 Mass spectrometry

Samples were analyzed using mass spectrometry as previously described in Chapter 4.

5.2.2 Cell proliferation

Proliferation was assayed as previously described in Chapter 4. Briefly, a thymidine incorporation assay was used to measure proliferation 25 hours after ligand stimulus. Error bars represent the standard deviation from 4 different biological replicates for each condition.

5.2.3 Cell migration

Migration was assayed as previously described in Chapters 2 and 4. Briefly, a wound healing analysis provided rates of wound closure. Error bars represent the 95% confidence intervals for the fit of the slope using linear regression.

5.2.4 Partial least squares regression (PLSR)

The PLSR model was generated using a SIMCA-P (10.0) software package as described elsewhere [10]. Briefly, a *MxN* data matrix (**X**) was generated from the mass spectrometry signaling dataset. Each column corresponded to one of the following four metrics: protein phosphorylation at 5 minutes, 10 minutes, 30 minutes, or the integral of protein phosphorylation from 0-30 minutes, used as a proxy for the total amount of protein phosphorylation during the 30 minute duration. There were 248 columns in

total, corresponding to the 4 metrics x 62 phosphosites = 248 columns. Each row represented a different cellular condition, with a total of six rows corresponding to parental serum-free, parental plus EGF (100 ng/ml), parental plus HRG (80ng/ml), 24H serum-free, 24H plus EGF (100 ng/ml), and 24H plus HRG (80 ng/ml). An *MxP* matrix (**Y**) was generated from the cellular output data, with the rows corresponding to the same cellular conditions listed above and the columns representing cell migration and cell proliferation. All data were mean-centered and scaled to unit variance.

PLSR was used to solve the regression problem:

$$\mathbf{Y = Xb + e} \tag{1}$$

where b is the vector containing the regression coefficients and e is the residual. A non-iterative partial least squares (NIPALS) algorithm was used [11, 12]. It is instructive to note that the NIPALS algorithm applied to a single *MxN* matrix (**X**) is the representation of the matrix as a sum of outer products such that:

$$\mathbf{X = t_1 p_1' + t_2 p_2' + \ldots + e} \tag{2}$$

where t_i is called the scores vector and represents one dimension in the orthogonal basis set for the column space and p_i is called the loadings vector and represents one dimension in the orthogonal basis set for the row space. Application of NIPALS in this way is analogous to the singular value decomposition (SVD) of the matrix such that:

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^{\mathrm{T}} \qquad (3)$$

where each scores vector is equivalent to a column of the product $U\Sigma$, and each loadings vector equivalent to a column of V.

The NIPALS algorithm was implemented as described elsewhere [12]. Briefly, an iterative process is used to define two vectors, *w* and *c*, that maximize the following term:

$$[\mathrm{Cov}(t,u)]^2 = [\mathrm{Cov}(Xw,Yc)]^2 \qquad (4)$$

where *t* and *u* are the scores vectors for the **X** and **Y** matrices respectively. Loadings vectors for **X** and **Y**, called *p* and *q* respectively, are also defined as:

$$\mathbf{p} = \mathbf{X}^{\mathrm{T}}\mathbf{t}/(\mathbf{t}^{\mathrm{T}}\mathbf{t});\ \mathbf{q} = \mathbf{Y}^{\mathrm{T}}\mathbf{u}/(\mathbf{u}^{\mathrm{T}}\mathbf{u}) \qquad (5)$$

The PLS regression vector *b'* is defined as:

$$\mathbf{b'} = \mathbf{u}^{\mathrm{T}}\mathbf{t}/(\mathbf{t}^{\mathrm{T}}\mathbf{t}) \qquad (6)$$

The set of vectors *t*,*u*,*w*, and *c* are associated with the maximum eigenvalues for various covariance matrices, and once defined, their contribution is removed from the **X** and **Y** matrices, leaving a residual matrix that can be further modeled with a new set of

t,u,w,c,p, and *q* vectors. The matrices corresponding to each of these vectors are defined as **T**,**U**,**W**,**C**,**P** and **Q**. The residual matrices are defined as:

$$\mathbf{X}_i = \mathbf{X}_{i-1} - \mathbf{t}_i\mathbf{p}_i^T,\ \mathbf{Y}_i = \mathbf{Y}_{i-1} - \mathbf{t}_i\mathbf{c}_i^T \quad (7)$$

Where the maximum value of *i* (referred to as A later for clarity) depends on the results of cross-validation. Weights were defined as w^*, which are calculated from *w* to relate to the original X-matrix (and not the residual as calculated above) as:

$$\mathbf{W}^* = \mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1} \quad (8)$$

Each model was tested for goodness of prediction (Q^2) using a leave-one-out cross validation approach [13]. Briefly, cross-validation is performed by omitting an observation from the model development and then using the model to predict the **Y**-matrix values for the withheld observation. This procedure is repeated until every observation has been kept out exactly once. The prediction error sum of squares (PRESS) is then calculated as:

$$\mathbf{PRESS} = \sum_i \sum_m Y_{i.m}^{measured} - Y_{i,m}^{predicted} \quad (9)$$

Q^2 is then calculated as:

$$\mathbf{Q^2} = \mathbf{1.0} - \prod_{a=1}^{A} (PRESS / SS)_a \quad (10)$$

where *a* refers to each individual principal component in the model and *SS* is the sum of squares explained by principal component *a*. Standard errors and confidence intervals (as evaluated for the models presented) can be derived from cross-validation as well [13].

To evaluate the scores plot transitions (Tables 1-3), the vector $\mathbf{T}_{j,1:A} - \mathbf{T}_{k,1:A}$ was evaluated to represent the transition from the cellular condition described by row *k* of **T** to the cellular condition represented in row *j* of **T**. The inner product of this vector and the weight of variable *m* ($\mathbf{W}^*_{m,1:A}$) was evaluated for all variables (*m*=1:248) and then ranked by magnitude.

To identify the signaling metrics most important for the overall model, a weighted sum of squares (also known as the variable importance for projection [VIP]) value for each variable (*k*) was calculated according to the following formula [13]:

$$VIP_k = \sqrt{\frac{K_T \sum_{a=1}^{A} w_{a,k}^2 SS_a}{\sum_{a=1}^{A} SS_a}} \qquad (11)$$

where K_T is the total number of variables and the rest of the variables are as described above. Signals having multiple metrics that ranked in the top 30 highest VIP scores were chosen for the reduced model.

To evaluate the importance of a given signal for an output, the inner product of each metric with a given output was evaluated by:

$$\mathbf{W}^*_{m,1:A} \bullet \mathbf{C}_{n,1:A} \tag{12}$$

for all variables (m=1:248) for each of both outputs (n=1:2). To calculate the unique contribution of the signaling metrics to a given output (say n = 1), we evaluated the following expression:

$$\mathbf{W}^*_{m,1:A} \bullet \mathbf{C}_{1,1:A} - \mathbf{W}^*_{m,1:A} \bullet \mathbf{C}_{2,1:A} \tag{13}$$

for all m. Overall contributions to the model were calculated as:

$$\mathbf{W}^*_{m,1:A} \bullet \mathbf{C}_{1,1:A} + \mathbf{W}^*_{m,1:A} \bullet \mathbf{C}_{2,1:A} \tag{14}$$

5.3 Results

5.3.1 Mass spectrometry approach measures 62 intracellular signals in human mammary epithelial cells

As previously described, we developed and employed a mass spectrometry approach to measure the effect of HER2 overexpression in 184A1 human mammary epithelial cells (HMEC) [10]. Two closely-related cell lines with known receptor expression levels were used; a parental cell line (P) with 200,000 EGFR, 20,000 HER2, and 20,000 HER3, and 24H cell line with 200,000 EGFR, 600,000 HER2, and 30,000

HER3 per cell [10]. Both cell lines have very low levels of HER4. Thus, the 24H cell line was used to assess effects of HER2 overexpression, with the parental cell line serving as a baseline for these measurements. HMEC's were treated with either saturating levels of EGF or HRG, and under each treatment the tyrosine phosphorylation of 62 phosphosites was quantified at 0, 5, 10, and 30 minutes. Figure 5-1 displays the 248 time courses collected. Our measurements revealed the dynamic activation of molecules both commonly associated with ErbB signaling (*e.g.*, extracellular regulated kinase 1 [ERK1] and SH2-containing protein [Shc]), and those less commonly associated with the ErbB network (*e.g.*, human transferrin receptor [TfR], ephrin A2 receptor [EphA2], and the previously unidentified KIAA 1217). Comparison with previously published maps of ErbB and migration-associated signaling networks reveals broad network coverage with the 62 measured signals (Figure 5-2, [7, 14]).

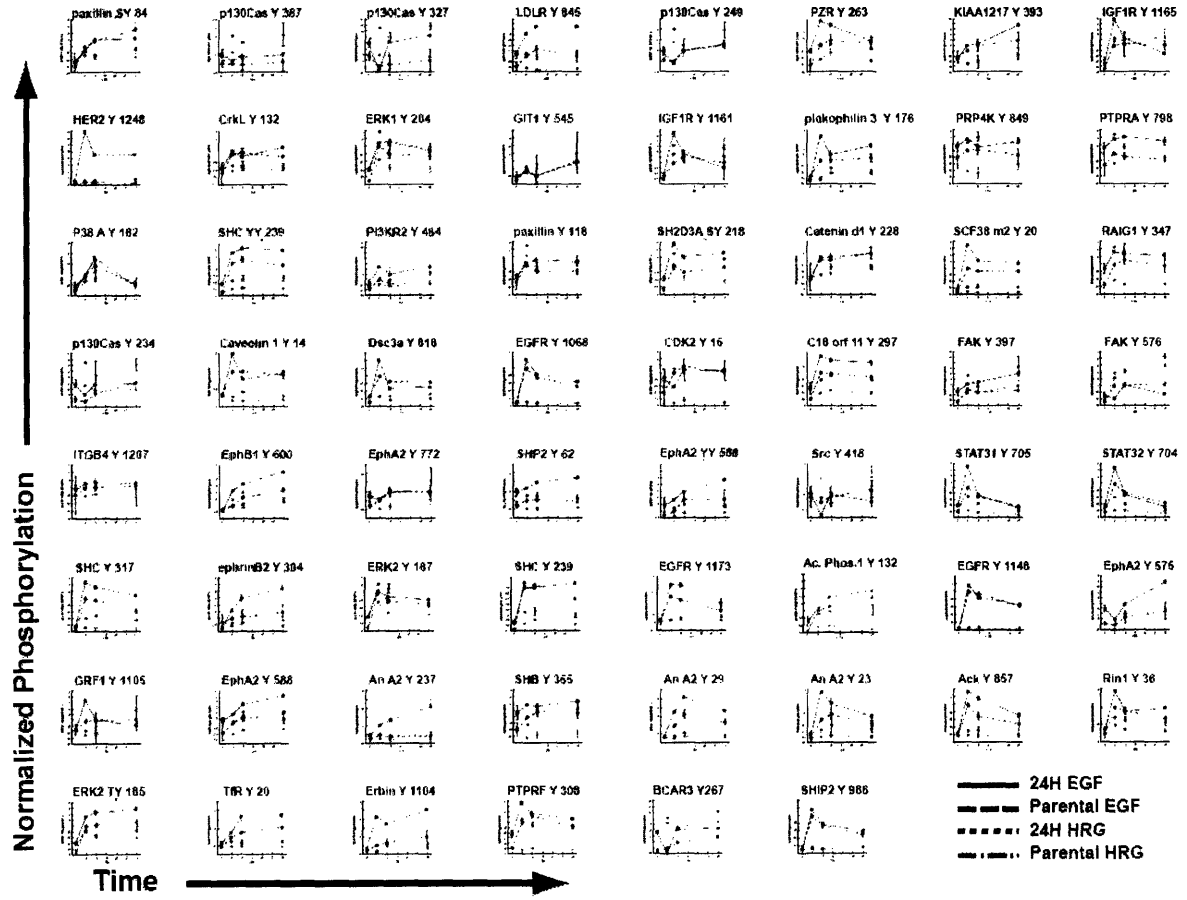


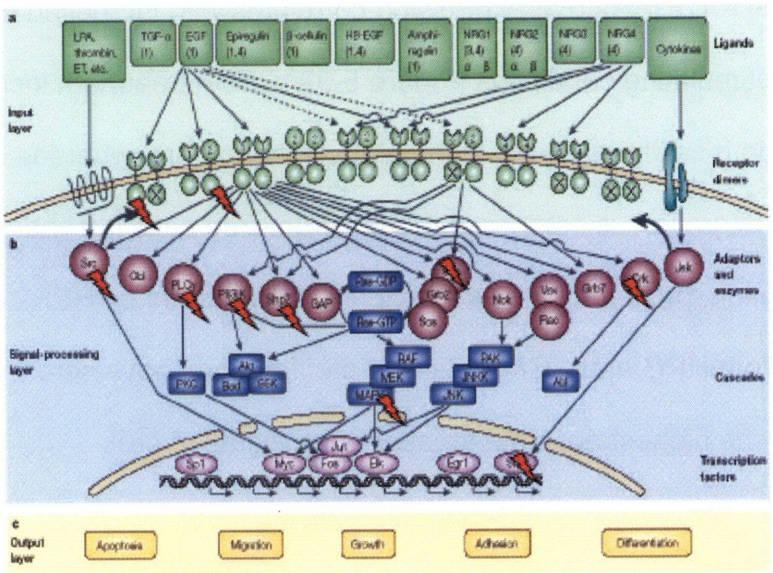
Figure 5-1: A mass spectrometry approach measured 248 phosphorylation profiles. The title at the top of each entry indicates the phosphosite measured. Median normalized phosphorylation (see Methods) is plotted at 0, 5, 10, and 30 minutes. For each phospho-site, four conditions were measured: parental + HRG (80 ng/ml), 24H + HRG (80 ng/ml), parental + EGF (100ng/ml), and 24H + EGF (100 ng/ml). Error bars indicate ± standard deviation. The data were obtained in [10].

Figure 5-2 (below): Measured signals provide broad coverage in ErbB- and cell migration-associated signaling networks. Multiple measured phosphotyrosine sites have been previously reported to be critical for ErbB-associated signaling (A) as well as for cell adhesion- and migration-associated signaling (B).

A.

Targeted (pTyr) Phosphoproteomics

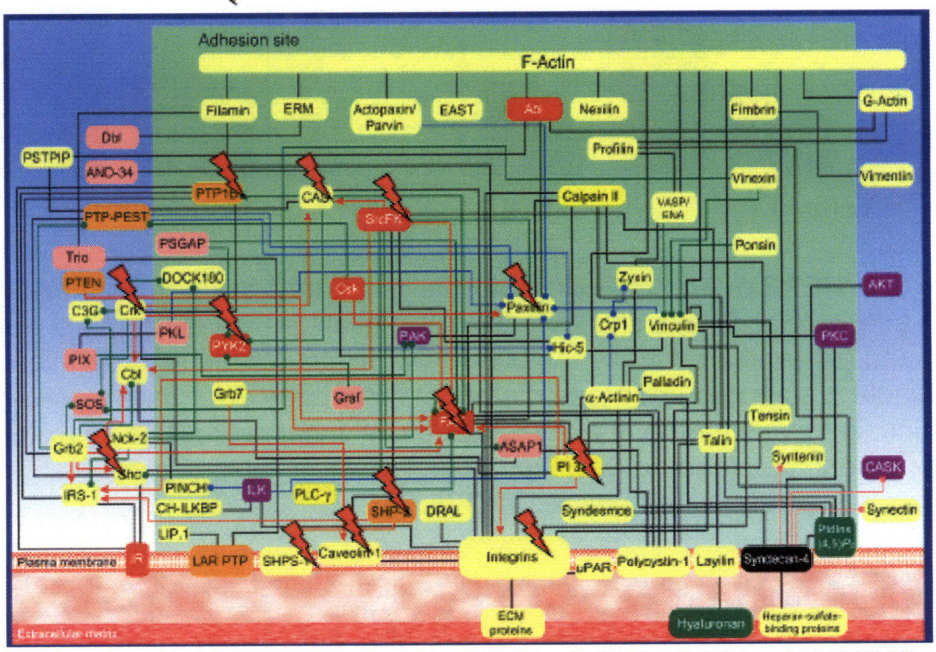
Phosphorylated tyrosine ( )mapped on mitogenesis-associated proteins



Yarden and Slikowski, *Nat Rev Mol Cell Biol.* 2001 Feb;2(2):127-37

B.

Phosphorylated tyrosine ( )mapped on cell migration-associated proteins



From Zamir E. & Geiger. B. (2001). J. Cell Science 114, 3577-3579

5.3.2 Cell proliferation and migration are differentially affected by HER2

Cellular migration and proliferation were measured under the same conditions described above [10]. HER2 overexpression correlates with increased cell migration under all ligand stimulating conditions (Figure 5-3A). EGF treatment increased the rate of migration for both cell lines by the highest absolute amount, whereas HRG treatment did not increase migration as compared to the serum-free case in either cell line. In contrast to migration, proliferation was not significantly altered by HER2 overexpression (Figure 5-3B). Both HRG and EGF increased proliferation above serum-free levels, with EGF stimulating the highest absolute amount of proliferation [10].

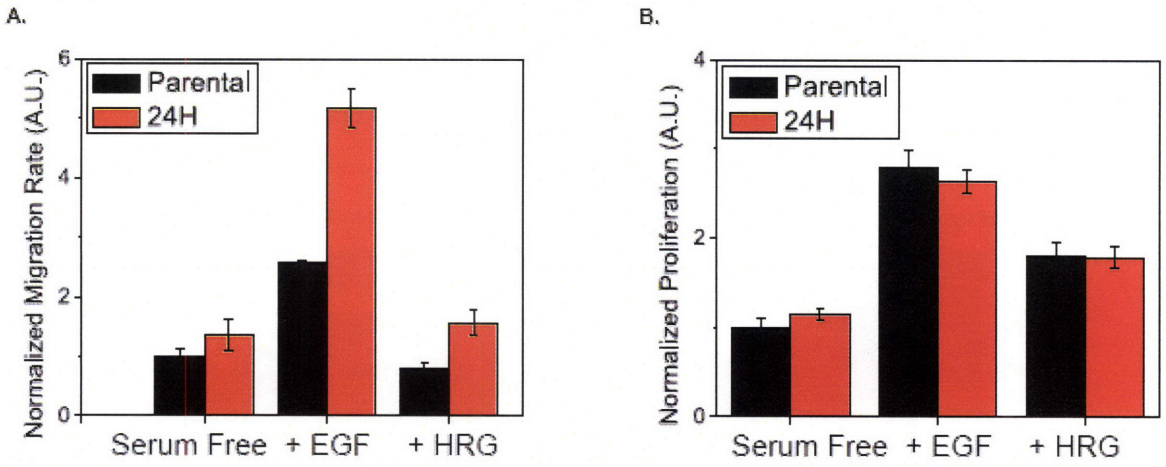


Figure 5-3: HER2 overexpression affects migration but not proliferation. Parental (black) and 24H (red) data shown for: (A) Migration, as measured by a wound healing assay, and (B) Proliferation, as measured by [3H] thymidine incorporation. All error bars denote ± standard error (see methods). Migration error bars represent the 95% confidence intervals for the fit of the slope using linear regression. Proliferation error bars represent the standard deviation from 4 different biological replicates for each condition. The data were obtained in [10].

5.3.3 Model analysis reveals phenotype-relevant signaling sets that characterize ligand and receptor expression transitions

A model based on partial least squares regression was created to linearly regress signaling metrics onto cellular migration and proliferation metrics ([10] and Materials and Methods). The model accurately recapitulated experimental data and had a goodness of prediction (Q^2) of 0.89 (Figure 5-4). Each signal comprised four metrics: the 5, 10, and 30 minute phosphorylation levels and the integral of the time course from 0 to 30 minutes. The integral was used as a measure for total phosphorylation. The zero minute time point was included in the row of serum-free observations (see Materials and Methods).

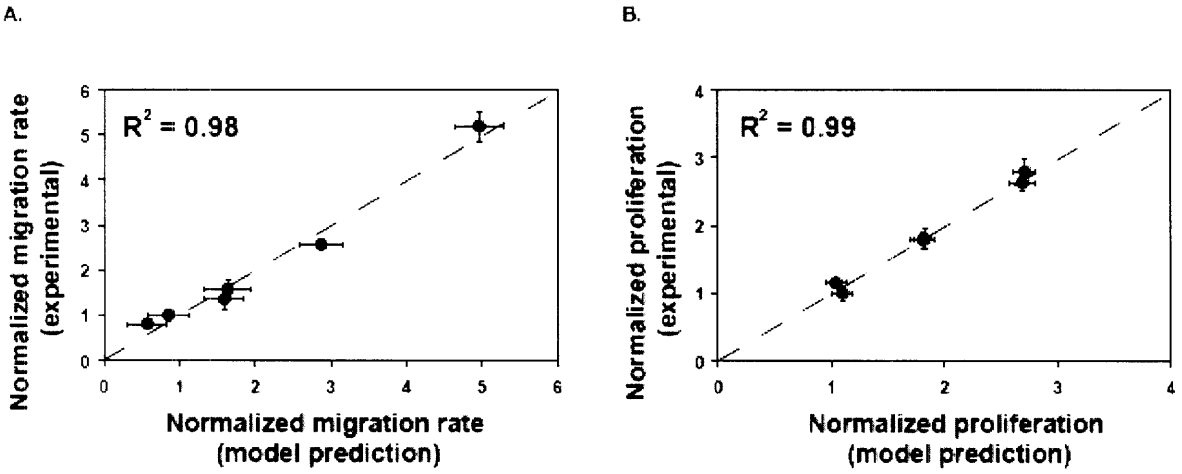


Figure 5-4: A linear model accurately recapitulates experimental data. Results from partial least squares regression (PLSR) modeling show that a computational model based on experimental data generates predicted values of cellular migration and proliferation that correlate strongly with experimentally measured values. Experimental values of migration (A) or proliferation (B) are graphed along the ordinate, and model predictions of migration (A) or proliferation (B) are given along the abscissa. R^2 values indicate the data's goodness of fit to the line y = x. Experimental error bars denote 95% confidence intervals for migration (see methods) and ± standard deviation for proliferation. All computational error bars are calculated from jack-knifing as a standard error.

Decomposition via PLSR of the signal and response datasets provided a reduced-dimension map (called the scores vector **t**, see Materials and Methods) on which network signaling changes in response to ligand or receptor perturbation could be interpreted. The plot axes (referred to as principal components) are linear combinations of the signaling metrics described above. Figure 5-5 plots the changes corresponding to ligand or receptor perturbation on a two dimensional graph whose axes are the first two principal components (the third component, which captures only 4% of the Y-block variance described by the full model, is omitted here for ease of visualization). The plot shows, as expected, that HRG treatment stimulates the signaling network distinctly from the EGF treatment case. If we had imagined that HRG treatment activated the same set of signals as EGF but at a different magnitude, then we would have expected to see the +EGF and +HRG vectors overlapping, with one being longer than the other. In reality we know that HRG activates different dimers than does EGF (*i.e.,* HER3-HER2 versus EGFR-HER2 or EGFR-EGFR), which in turn drives the differential activation of the signaling network. Interestingly, the difference between EGF and HRG signaling is larger in 24H cells relative to parental cells (77 degrees versus 37 degrees), due in large part to a drastic shift in EGF-induced signaling with HER2 overexpression (Figure 5-5B). If we take a linear superposition of changes in signaling due to HER2 overexpression in the absence of ligand (*i.e.*, the signaling changes under serum-free conditions between P and 24H cells) and the changes in signals we expect as we add a ligand (*i.e.*, signaling changes as either HRG or EGF are added to P cells), we can approximate the signals generated by 24H cells under HRG treatment (Figure 5-5C). We cannot do the same, however, for 24H cells under EGF treatment, emphasizing the

non-additive interplay between changes in cellular ligand-receptor conditions and the signals they generate (Figure 5-5C). In the case of our HMEC system, previous quantitative measurement and modeling of dimer kinetics has shown that HER2 overexpression in the presence of HRG leads to a relatively small shift in the number of HER2-HER3 dimers (~10,000). Alternatively, HER2 overexpression in the presence of EGF leads to a large increase in EGFR-HER2 dimers (~150,000) and a decrease in EGFR homodimers (~65,000) [10]. Changes in signaling with HER2 overexpression under serum-free conditions could be due to either spontaneous homodimerization or autocrine signaling. Thus, given our analysis of the scores plot, we hypothesize that increases in HER2 under HRG treatment principally add to the signaling network through the independent addition of signals generated through HER2 homodimerization or autocrine signaling. HER2 overexpression under EGF treatment, however, results in the addition of these signals plus a novel suite of signals generated primarily through an increase in EGFR-HER2 dimers and loss of EGFR homodimers (see Appendix 5).

Although the scores plots allow us to visualize signaling changes, it is often of interest to relate observed differences back to original measured signaling metrics. We accomplish this by taking the inner product of any vector in the scores plot with the principal component axes, and thereby derive lists of proteins that most strongly correlate with the transition associated with the vector (see Materials and Methods). Figure 5-6 outlines this approach and indicates the 30 cell state transitions analyzed. We discuss here a subset of those transitions, but all results are available in Supplementary Materials of [15].

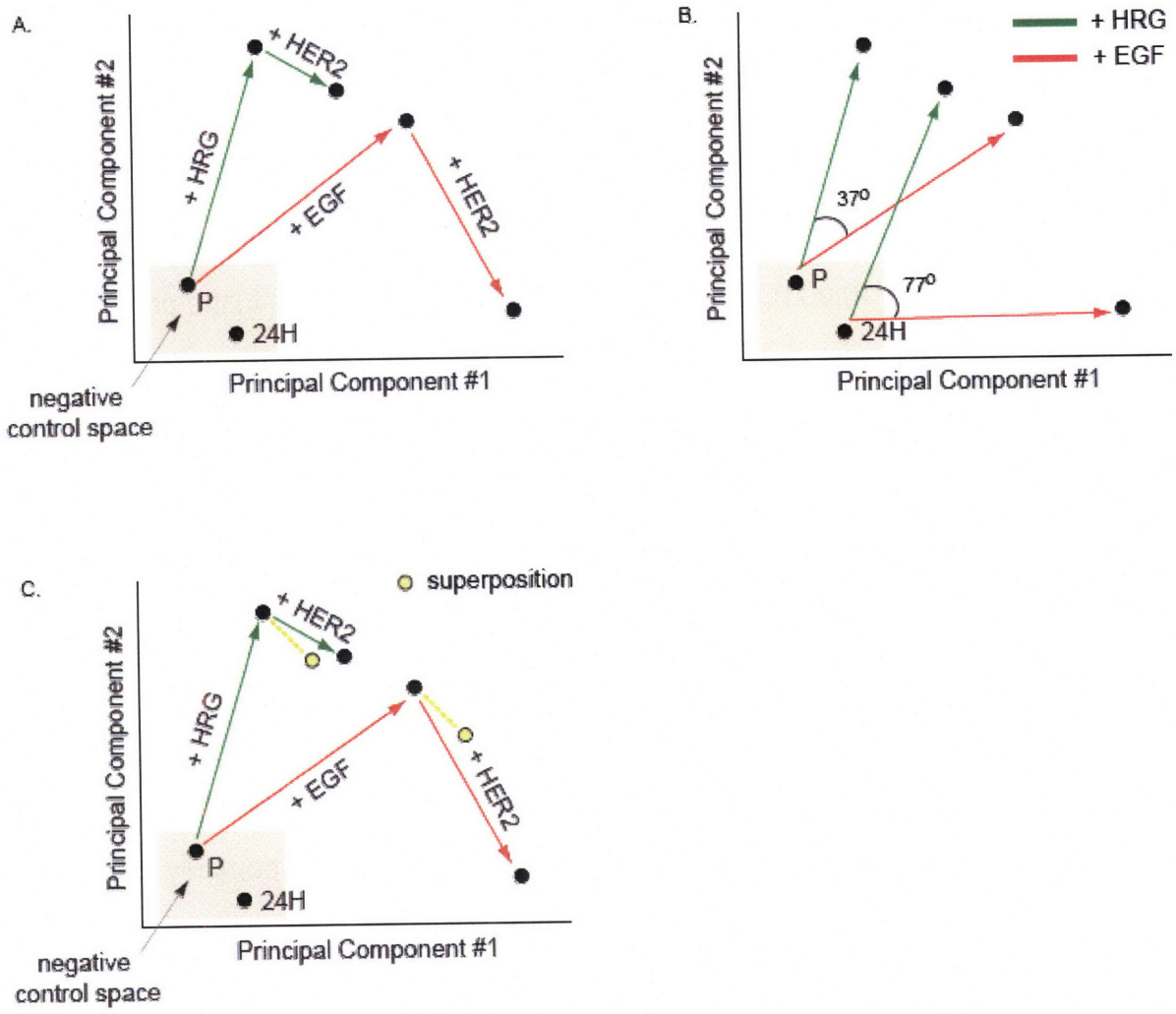


Figure 5-5: PLSR-generated scores plot reveals different signaling strategies for EGF and HRG. A scores plot identifies separation in signaling strategies associated with receptor overexpression or differential ligand treatment along two signaling axes (A). HRG and EGF treatment give rise to different sets of signals, and the difference is exaggerated in 24H cells (B). The linear superposition of the difference vector between 24H and parental serum-free conditions and each ligand's vector explains 24H+HRG signaling but not 24H+EGF signaling (C).

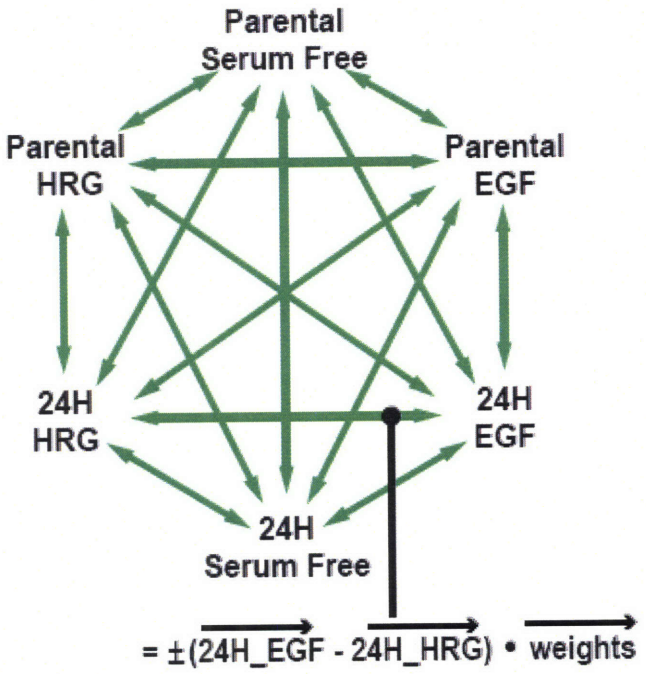


Figure 5-6: A strategy for the study of 30 possible cellular transitions. Each arrow represents the difference vector between two cell states. The changes in the signaling network associated with the transition between the two states are calculated by taking the inner product of the difference vector with the weights vector (see Materials and Methods).

HER2 overexpression in the presence of EGF, as discussed above, produced interesting signal network changes and increased cell migration but did not affect cell proliferation (see Figure 5-3). Using our approach, we derived proteins positively correlating most strongly with HER2-associated increases in motility under EGF treatment (Table 5-1A). HER2 phosphorylation at tyrosine 1248 and that of Crk-associated substrate (p130Cas) at tyrosine 327 feature prominently in the list, agreeing with previous reports linking the HER2-specific activation of p130Cas to increased invasion in breast epithelial cells [9]. Another protein listed is annexin A2 (also known as lipocortin 2), a molecule previously known to mediate cytoskeletal-membrane interactions, therefore linking it to critical processes governing cell migration [16]. While

expression of annexin A2 has been found to decrease or increase cell migration, depending on the cell system, little is known about how its phosphorylation correlates with cell migration [17, 18]. Here, we speculate that annexin A2 is part of the mechanism that increases cell migration under HER2 overexpression in the presence of EGF, and we identify a novel phoshorylation site, tyrosine 237, that may regulate its role in the activation of migration. Phosphorylation of SH2 domain-containing phosphatase SHP-2, another molecule that appears in Table 5-1A multiple times, has been shown to increase cell migration in breast cancer cells, although connection to particular phosphorylation events has been sparse [19]. Here we implicate the tyrosine 62 site in SHP-2's activation and eventual effect on cell migration. Interestingly, SHP-2 and annexin A2 have been found to complex in endothelial cells, suggesting the possible presence of a co-regulatory scheme in our HMEC system [20].

A list of proteins most negatively correlated with phenotype includes all measurements of EGFR phosphorylated at tyrosine 1173 as well as Src, which has been shown to phosphorylate EGFR tyrosine 1173 (Table 5-1B, [21]. These molecules exhibit decreased phosphorylation in response to increasing HER2 expression and concomitant increases in migration under EGF treatment. The tyrosine 1173 site on EGFR has been shown to recruit SH2 domain-containing phosphatase SHP-1, which helps coordinate EGFR dephosphorylation and mitogen-activated protein kinase (MAPK) deactivation [22]. We speculate that decreased tyrosine 1173 phosphorylation is part of the mechanism through which HER2 increases the downstream signaling governing increased migration.

Next, we sought to understand the effect of changing ligand under given receptor expression levels (Table 5-2). Substitution of EGF for HRG in 24H cells increases both proliferation and migration, although the absolute increase in migration is greater (Figure 5-3). Interestingly, many proteins that negatively correlate with this transition are linked to migration-relevant p130Cas pathway (Table 5-2B). This pathway includes Src and its substrates focal adhesion kinase (FAK) and p130 Cas (FAK also phosphorylates p130Cas) [23, 24]. Counterintuitively, then, EGF appears to negatively regulate a classical migration pathway to increase cell migration. The new EGF-stimulated signals not only increase migration but proliferation as well. In Table 5-2A, we observed proteins previously linked in the literature to migration (*e.g.*, annexin A2, glucocorticoid receptor DNA binding factor 1[GRF1]), and others linked to proliferation (EGFR, desmocollin-3 [Dsc3a) [7, 10, 16]. KIAA 1217 is a previously unidentified protein that warrants further investigation for its potential role in EGF-mediated proliferation and migration.

As mentioned above, the signaling changes associated with increased HER2 expression in the presence of HRG are very similar to the same change under serum-free conditions. In both cases, HER2 overexpression leads to an increase in migration but not proliferation. Signaling metrics that positively correlate with this transition include p130Cas and FAK, indicating that increased migration may be mediated through this migration-associated pathway (Table 5-3A and Supplementary Table 1 in [15]). Serine/threonine protein kinase PRP4 homolog (PRP4K) and protein tyrosine phosphatase receptor type A (PTPRA) are two additional molecules that correlate with the increased migration suggesting a novel role for both in HER2-mediated migration.

Activation of these molecules, as mentioned above, may be due to spontaneous HER2 homodimerization or autocrine signaling. Considering Table 5-3B, we note that EGFR tyrosine 1173 negatively correlates with the increase in migration, just as we observed for HER2 overexpression in the presence of EGF. HRG treatment, however, is generally not thought to regulate EGFR phosphorylation, and the presence of tyrosine 1173 in Table 5-3B suggests that HER2 overexpression in the presence or absence of HRG drives migration through autocrine signaling that activates EGFR-HER2 heterodimers. Indeed, we have shown that there is a measurable but low amount of transforming growth factor alpha (TGF-α) produced by both the parental and 24H cells (personal communication, Lisa Joslin and Douglas Lauffenburger), thus pointing to a potential mechanism for autocrine-induced signaling changes.

5.3.4 A nine-signal reduced model recapitulates full model performance

A reduced model based on a fraction of the 62 originally measured phosphorylation sites would be useful for the future study of HER2 effects when full network measurement is not possible. We identified nine phosphorylation sites on six proteins that recapitulated the performance of the full model. We refer to this subset of signals as a 'network gauge': a small number of phosphorylation sites that together can be interrogated to predict levels of proliferation and migration. To rank phosphorylation events according to their importance in the full model, we used a weighted sum of squares technique (see Materials and Methods). Phosphorylation sites whose metrics appeared more than once in a list of the top 30 weighted sum of squares metrics were

selected for the reduced model. Using this method, phosphorylation sites on the following six proteins were identified as important: transferrin receptor (TfR), annexin A2, activated cdc42-associated kinase (ACK), SH2-containing inositol polyphosphate 5-phosphatase (SHIP-2), SH2-containing protein (Shc), and solute carrier protein 38 (SCF38, also known as SNAT2 or ATA2). All measured tyrosine phosphorylation sites were included for these molecules except for the tyrosine 237 site on annexin A2, since it was not represented at any time on the top 30 list. A model generated with this six protein set had an excellent goodness of fit to experimental data compared to the full model (Figure 5-7). In addition, the reduced model maintained a high goodness of prediction (Q^2 = 0.95). A model based on the bottom 30 ranked metrics had a goodness of fit less than 0.30 to experimental data, suggesting that our ranking appropriately isolated highly informative sets of phosphorylation metrics.

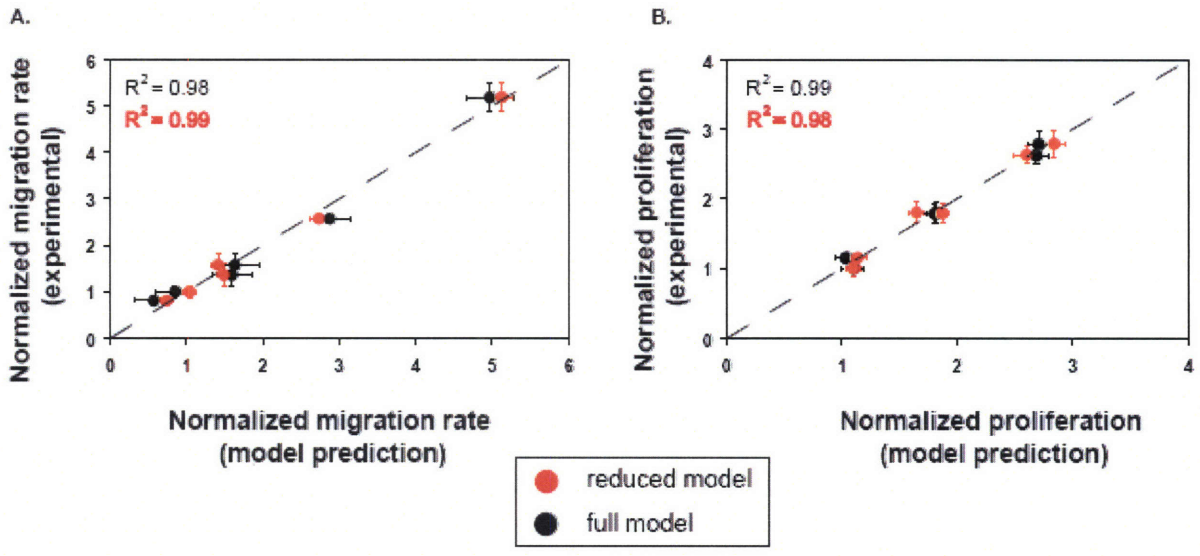


Figure 5-7: A linear model based on nine of the original 62 signals recapitulates experimental data and matches full model performance. Results from partial least squares regression (PLSR) modeling show that a computational model based on six signals generates predicted values of cellular migration and proliferation that correlate strongly with those predicted by the full model and experimentally measured values. Experimental values of migration (A) or proliferation (B) are graphed along the ordinate,

and full-model predictions (black) and reduced-model predictions (red) of migration (A) or proliferation (B) are given along the abscissa. R^2 values indicate the data's goodness of fit to the line y = x. Experimental error bars denote 95% confidence intervals for migration (see methods) and ± standard deviation for proliferation. All computational error bars are calculated from jack-knifing as a standard error.

The somewhat surprising makeup of the network gauge prompted further investigation as to why these particular six proteins were so informative. Shc's tyrosine site at 239/240 regulates c-myc activation, its tyrosine site at 317 regulates MAPK activation, and its phosphotyrosine-binding domain (PTB) domain is known to associate with phosphatidylinositol-3,4,5-trisphosphate (PIP3), although it is not known how Shc's binding affinity for PIP3 changes with tyrosine phosphorylation [25]. Thus, tyrosine phosphorylation of Shc affects multiple important signaling pathways leading us to speculate that the dynamic and quantitative measurement of Shc at both tyrosine phosphosites may serve as a sensitive indicator for the ultimate activation of pathways important for proliferation and migration. The afore mentioned annexin A2, a target of kinases such as Src and protein kinase C (PKC), has been found to mediate membrane-cytoskeleton interactions and its knockdown has been linked to decreased invasiveness in human glioma cell lines [26]. Annexin A2 clusters strongly with migration in the reduced model, revealing its role as a relatively migration-specific predictor. TfR endocytosis brings iron into the cell and is stimulated by tyrosine 20 phosphorylation [27]. Introduction of iron into the cell can be a major regulator of cell proliferation and growth, and TfR has also been linked to migration in endothelial cells [28, 29]. Moreover, TfR has been shown to partially traffic with EGFR, although it may internalize via a different mechanism [30-32]. We report here, to our knowledge, the first report of EGF-stimulated TfR phosphorylation and regulation via HER2 expression,

and we hypothesize that tyrosine phosphorylation of TfR is a reliable indicator of endocytic regulation in our system. ACK, downstream of cdc42, has been shown to bind clathrin and to regulate TfR trafficking, implicating it also as a part of the endocytosis readout [33]. ACK has additionally been shown to be an early transducer of EGFR signaling and to negatively regulate cell adhesion through the CrkII-p130Cas pathway, a pathway shown above to be important in HER2-mediated migration [9, 34, 35]. SHIP-2 is a phosphatase that acts on PIP3, and has been found associated with filamin and actin, implicating it directly in the regulation of both PI3K signaling pathways and cell migration [36]. Additionally, SHIP-2 regulates EGFR trafficking and associates with Shc in response to EGF binding, further suggesting that SHIP-2 phosphorylation is a sensitive readout for a wide variety of different signaling responses [37, 38]. SCF38 is a System A amino acid transporter that responds to growth factor signaling [39]. Insulin stimulation leads to the recruitment of SCF38 to the cell membrane through a PI3K-dependent signaling mechanism that coordinates its trafficking from the endosomal compartments [40]. Little is known about potential roles for SCF38 in ErbB signaling, but based on our model results, further biochemical studies are warranted.

Two interesting themes, then, emerge from the network gauge findings: (a) the inclusion of a group of molecules linked to endocytosis, namely TfR, ACK, and SHIP-2; and (b) the high proportion of molecules known to interact with PI3K or PIP3, namely Shc, SHIP-2, TfR, and SCF38. We will elaborate further on these themes in the Discussion section.

5.3.5 Models based on parental cell data alone accurately predict the effects of HER2 overexpression on proliferation and migration

We investigated whether the full model and the network gauge trained only on parental cell data could predict migration and proliferation levels in response to HER2 overexpression. We trained models on parental serum-free, EGF, and HRG data, performed PLSR to calculate regression coefficients, and then used the measured 24H signal values in the regression equation to predict proliferation and migration. We found that both the full 62-signal model and the network gauge were able to predict proliferation and migration in 24H cells ($R \geq 0.99$, Figure 5-8, Figure 5-9).

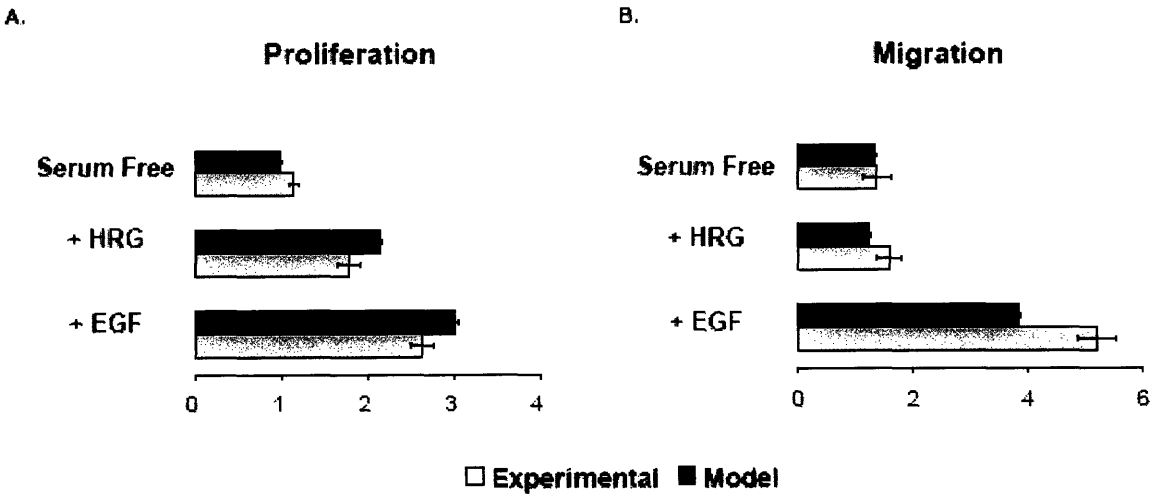


Figure 5-8: A reduced-signal PLSR model based only on parental cell data predicts 24H proliferation and migration. A PLSR model of nine signals constructed from parental cell data only was used to predict (A) proliferation and (B) migration levels in 24H cells and then compared to measured experimental values. Experimental error is denoted by 95% confidence intervals for cell migration (see methods) and ± standard deviation for cell proliferation. Computational error bars were calculated as the 95% confidence intervals based on jack-knifing.

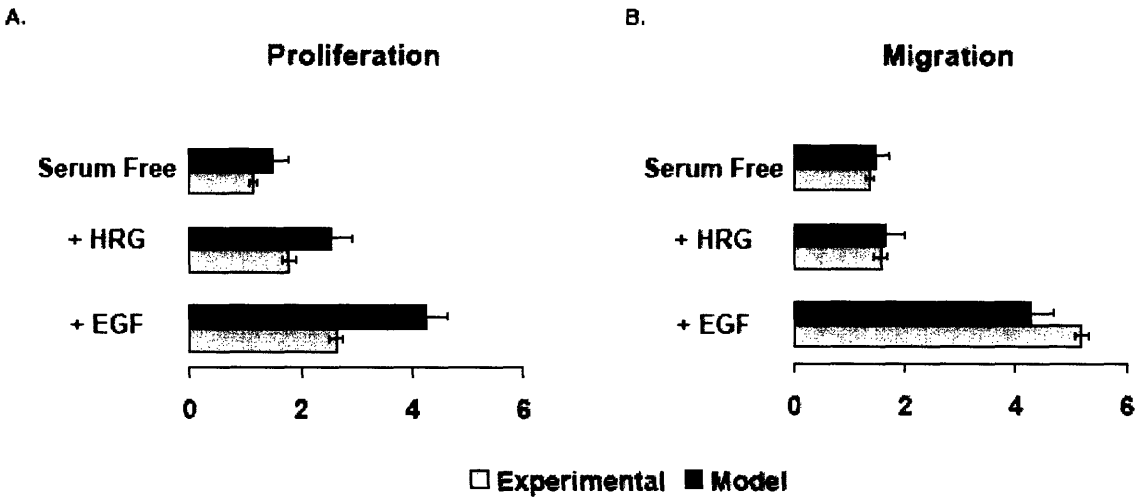


Figure 5-9: A linear model based only on parental cell data predicts 24H proliferation and migration. A PLSR model constructed from parental cell data only was used to predict (A) proliferation and (B) migration levels in 24H cells and then compared to measured experimental values. Experimental error is denoted by 95% confidence intervals for cell migration (see methods) and ± standard deviation for cell proliferation. Computational error bars were calculated as the 95% confidence intervals based on jack-knifing.

Although the model does not match the migration or proliferation results exactly in terms of absolute quantities, it captures salient trends such as the sharp increase in migration under EGF treatment for 24H cells and the trend of increasing proliferation when HRG is substituted with EGF. Interestingly, the model also predicts a slightly decreased amount of migration under HRG stimulation in comparison to the serum-free condition, which corresponds to the lack of HRG-mediated migration observed experimentally. The predictive capability of the model indicates that although HER2 overexpression changes intracellular signals drastically, the rules by which those signals are brought together to affect proliferation and migration, as defined by PLSR, remain the same as in the parental cells.

5.3.6 PLSR analysis reveals signals that uniquely correlate with migration and proliferation

Identification of molecules highly but uniquely associated with either proliferation or migration are of value when considering strategies to knock-down one behavior without affecting the other. We reported previously the top 20 signals associated with migration and proliferation through an analysis of reduced-dimension PLSR plots [10]. Interestingly, however, we noted that most signals having a high projection for one output have a non-zero projection for the second output. Thus, the importance of a signal with respect to cellular phenotype is a quantitative assessment, and we can derive behavior-specific protein metrics by ranking the inner product of the metric and the relevant behavior after the inner product of that same metric with the second behavior has been subtracted (Materials and Methods). Tables 4A and 4B list the phosphorylation sites that most strongly and uniquely correlate with migration and proliferation respectively. Table 5-4A indicates that attractive migration-specific targets include annexin A2 237 and HER2 1248 tyrosine phosphorylation. Likewise, Table 5-4B indicates that Dsc3a or catenin-δ1 are good targets for proliferation-specific inhibition. If the goal is to perturb one behavior without perturbing the other at all, then the corresponding protein targets could be calculated by making a list of proteins that had close to zero projection along the output one desired not to affect. The problem with this strategy is that the projection along the behavior one desires to affect may be small as well.

Table 5-4C lists those proteins that positively correlate with both outputs to the greatest degree (*i.e.*, the sum of the inner products with both migration and

proliferation). While there exists (as expected) some redundancy with the migration and proliferation lists we published earlier ([10]), novel proteins such as p130Cas, FAK and PRP4K have an elevated importance when their effects on migration and proliferation are summed. Although p130Cas is usually associated with control of invasion and cell motility in HER2 overexpressing breast epithelial cells, recent reports link it to proliferative control in HER2-dependent mammary epithelial tumorigenesis [41]. FAK phosphorylated at the tyrosine 576 site also appears on the list while being absent from the individual migration or proliferation lists. FAK regulates breast cancer cell migration and invasion, and also plays a role in cell cycle and proliferative control, although evidence for this latter role has been sparse in HER2 overexpressing systems [42-44]. Here, we hypothesize that FAK plays a critical role in the control of both proliferation and migration in the HMEC cell line, and taken together with the appearance of p130Cas and Src, we hypothesize that the well-described Src-FAK-Cas pathways play an essential role in the control of both proliferation and migration in HMEC cells. This activity could be captured by our network gauge in numerous ways, for instance information about Src activity may be included via annexin A2 phosphorylation levels.

Table 5-4D lists the 20 proteins that demonstrated weakest correlation with both migration and proliferation (*i.e.*, the sum of their projections along both behaviors were small). The list is of some interest for two reasons: (a) some of the proteins are also listed in Tables 4A-C, but at different times or at different phosphorylation sites (*i.e.,* FAK or p130Cas) indicating the need for phospho-specific, time-resolved data; and (b) the presence of some proteins presumed to be of substantial relevance to HER2-associated signaling, such as extracellular signal-regulated kinase (ERK). The reason

ERK is not important for the model here is that its variance upon receptor or ligand perturbation does not consistently linearly correlate with consequent proliferation or migration measures across all conditions. Nonetheless, we have observed that treatment with mitogen-activated protein extracellular kinase (MEK) inhibitor has an effect on the proliferation of HMEC cells (PD98059, unpublished observations). This raises an important caveat concerning the model -- that there can be proteins important for cell responses that do not consistently correlate across all or most treatment conditions. We do not claim to be able to identify all of the false negatives (or positives), but rather state that a model of this type remains predictive and enabling of conceptual insights.

5.4 Discussion and conclusions

We have demonstrated the use of PLSR to characterize the relative importance of tyrosine phosphoryation events to cell migration and proliferation in a human mammary epithelial cell line with varying HER2 expression levels under both EGF and HRG treatment. In addition, we have identified an important subset of molecules from our original large signaling dataset to serve as a network gauge for the prediction of migration and proliferation (Figure 5-10). Our results both highlight previously identified elements in the HER2 signaling network, and suggest new pathways and targets critically implicated in HER2-mediated signaling and its effect on migration and proliferation.

Scores plot analysis (Figure 5-5) helped generate global intuition as to how different combinations of ligand and receptor expression activated the phosphotyrosine signaling network. We related these changes back to original measurements through the use of inner products, generating lists of proteins correlated with any given ligand or receptor transition. Because the lists are derived after applying PLSR, the proteins highlighted have already been identified as important for the description of changes in cellular behavior. This procedure represents an improvement over traditional analysis of large mass spectrometry datasets (usually fold-change analysis) and demonstrates, to our knowledge, the first time an approach based on inner products has been used to extract understanding from PLSR-based biological models. Our lists (Tables 1-3) show that a particular behavior may be controlled through different network signaling strategies depending on cellular input. For instance, when EGF replaces HRG in 24H cells, migration is stimulated through a different set of molecules than are used to elevate migration when HER2 levels are increased.

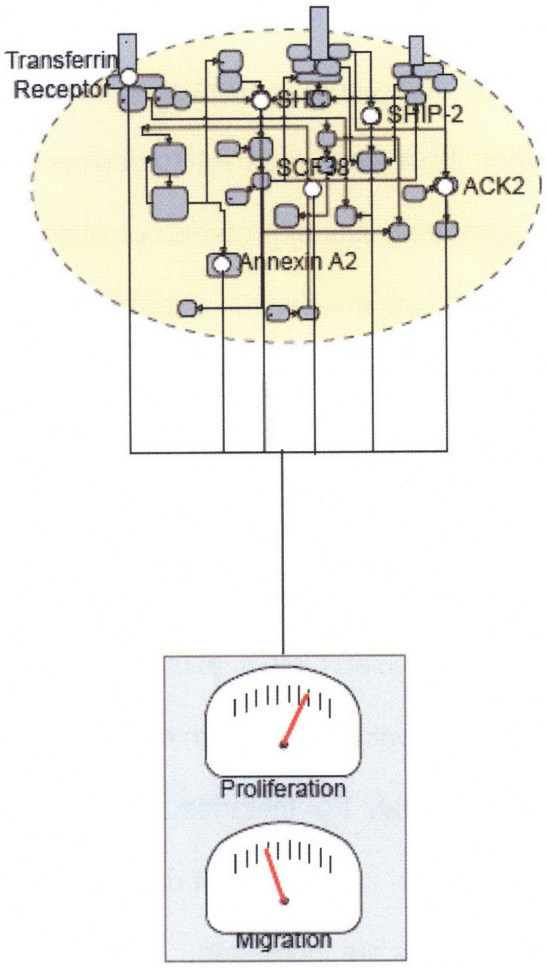


Figure 5-10: A network gauge predicts cell behavior and suggests critical elements of network architecture. A nine-signal PLSR model (A) that reliably predicts proliferation and migration includes transferrin receptor (TfR), annexin A2, solute carrier protein 38 (SCF38), SH2-containing protein (Shc), SH2-containing inositol polyphosphate 5-phosphatase (SHIP-2), and activated cdc42-associated kinase (ACK).

The reduction of the mass spectrometry dataset to nine highly informative phosphorylation sites on six proteins suggests elements of network architecture that likely control migration and proliferation, namely endocytosis and signaling through PIP3- and PI3K-mediated pathways. Three of the six highly informative proteins, TfR, SHIP-2, and ACK, are all linked to endocytosis [27, 33, 38]. The tight connection

between endocytic regulation and the signaling networks governing cell migration and proliferation has been documented, most powerfully in a recent study using RNA interference against the human kinome [45]. The results of this study indicate that more kinases than previously appreciated are involved in endocytosis, and taken together with other recent efforts, implicate endocytosis as a high-level regulator and sensor of cell signaling networks [45, 46]. Endocytosis can occur via many different mechanisms, principally clarthrin-mediated endocytosis (CME) and caveolar/raft-mediated endocytosis (RCE), with each mechanism regulating different sets of kinases and cell behaviors [45, 46]. The fact that TfR endocytosis was identified as highly informative instead of EGFR endocytosis might be due to the fact that EGFR internalization is mediated by both CME and RCE after treatment with high amounts of EGF, whereas TfR is thought to internalize independent from RCE [31]. The dynamic and quantitative resolution in our signaling assay was most likely critical for the capture of endocytic events, as endocytosis strongly regulates both signal duration and intensity. Furthermore, although our assay did not measure spatial distribution, endocytic information may have served as a proxy for that, further explaining its presence in the reduced model. Signaling through PI3K and PIP3 affects both commonly recognized downstream targets, such as protein kinase B (PKB or AKT), and important distinct pathways such as those containing ERK and p53 [47]. A recent mapping of the complete ErbB signaling network reveals PIP3 and its upstream kinase PI3K as highly informative nodes upon which a large fraction of signaling information converges [48]. Not surprisingly, then, we identify four proteins in our network gauge that interact with or are downstream of PIP3 or PI3K. These molecules are: Shc, SHIP-2, TfR, and SCF38

[25, 40, 49]. Thus, model reduction not only identifies a network gauge, but also suggests salient elements of the signaling network.

The PLSR model's ability to predict levels of proliferation and migration in 24H cells given only data from parental cells indicates that although signals drastically change as we move from parental to 24H cells, the cell decides upon levels of migration and proliferation according to the same 'rules'. These rules are non-intuitive but amount to the calculation of behavior according to the regression equation given by the PLSR model. Identification of conserved algorithms used to control behavior across cell type highlights the potential to predict *a priori* how changes in signaling will affect cell behavior and gives insight into conserved themes for cellular decision making processes. Thus, the linear mapping of phospho-proteomic data onto cellular phenotype identified a key set of signals descriptive and predictive of phenotype in breast epithelial cells. It also identified subsets of signals that govern phenotype under either ligand or receptor perturbation, and in that process revealed new hypotheses about HER2-mediated signaling events. Of course, these hypotheses need to be tested through further focused molecular and biochemical work. Nevertheless, the modeling approach we introduce here is a powerful first step toward understanding signaling networks and the behaviors they control.

5.5 Tables

Table 5-1: Signaling metrics most important for changes in cellular response due to changes in HER2 expression under EGF stimulation. Analyzed results from PLSR generated scores plot reveals (A) the 20 signaling metrics most positively correlated with changes in cell behavior as HER2 levels are increased, and (B) the 20

signaling metrics most negatively correlated with changes in cell behavior as HER2 levels are increased.

Table 5-2: Signaling metrics most important for changes in cellular response due to varying ligand exposure in cells with high HER2 expression. Analyzed results from PLSR generated scores plot reveals (A) the 20 signaling metrics most positively correlated with changes in 24H cell behavior when HRG stimulation (80 ng/ml) is substituted with EGF stimulation (100 ng/ml), and (B) the 20 signaling metrics most negatively correlated with changes in 24H cell behavior when HRG stimulation (80 ng/ml) is substituted with EGF stimulation (100 ng/ml).

Table 5-3: Signaling metrics most important for changes in cellular response due to changes in HER2 expression under HRG stimulation. Analyzed results from PLSR generated scores plot reveals (A) the 20 signaling metrics most positively correlated with changes in cell behavior as HER2 levels are increased, and (B) the 20 signaling metrics most negatively correlated with changes in cell behavior as HER2 levels are increased

Table 5-4: Analysis of reduced dimension PLSR mapping produces signaling metric hierarchies for cell behavior. Analyzed results from the X-Y loadings plot produced through PLSR analysis reveal (A) the 20 signaling metrics uniquely correlated with migration, (B) the 20 signaling metrics uniquely correlated with proliferation, (C) the 20 signaling metrics correlated most strongly with both migration and proliferation, and (D) the 20 least correlated signaling metrics for both migration and proliferation.

Table 5-1A

Phosphorylation Site (P→24H, EGF)	Measurement Type
p130Cas Y 327	10 minutes
p130Cas Y 327	30 minutes
SHP-2 Y 62	30 minutes
SHP-2 Y 62	10 minutes
PRP4K Y 849	30 minutes
EphA2 Y/Y 588/594	5 minutes
SHP-2 Y 62	5 minutes
HER2 Y 1248	30 minutes
An A2 Y 237	30 minutes
HER2 Y 1248	10 minutes
LDLR Y 845	30 minutes
HER2 Y 1248	5 minutes
EphA2 Y 575	10 minutes
LDLR Y 845	5 minutes
PTPRA Y 798	30 minutes
PTPRA Y 798	10 minutes
An A2 Y 237	10 minutes
PRP4K Y 849	5 minutes
HER2 Y 1248	integral
An A2 Y 237	5 minutes

Table 5-1B

Phosphorylation Site (24H→P, EGF)	Measurement Type
paxillin S/Y 84/88	30 minutes
IGF1R Y 1161	integral
IGF1R Y 1165	integral
EGFR Y 1173	5 minutes
Dsc3a Y 818	integral
paxillin S/Y 84/88	5 minutes
FAK Y 576	integral
EGFR Y 1173	30 minutes
EGFR Y 1173	integral
EGFR Y 1173	10 minutes
FAK Y 576	30 minutes
Dsc3a Y 818	30 minutes
Src Y 418	integral
Dsc3a Y 818	10 minutes
GIT1 Y 545	5 minutes
IGF1R Y 1161	30 minutes
FAK Y 576	5 minutes
IGF1R Y 1165	30 minutes
Src Y 418	5 minutes
Src Y 418	30 minutes

Table 5-2A

Phosphorylation Site (HRG→EGF, 24H)	Measurement Type
GRF1 Y 1105	10 minutes
KIAA1217 Y 393	5 minutes
KIAA1217 Y 393	10 minutes
ITGB4 Y 1207	30 minutes
Caveolin 1 Y 14	5 minutes
EGFR Y 1068	30 minutes
plakophilin 3 Y 176	5 minutes
Ack Y 857	30 minutes
Caveolin 1 Y 14	30 minutes
EGFR Y 1068	10 minutes
GRF1 Y 1105	5 minutes
An A2 Y 29	30 minutes
EGFR Y 1068	5 minutes
An A2 Y 29	5 minutes
SHIP-2 Y 986	5 minutes
EGFR Y 1148	5 minutes
SHIP-2 Y 986	10 minutes
Dsc3a Y 818	5 minutes
ephrin-B2 Y 304	5 minutes
Ack Y 857	5 minutes

Table 5-2B

Phosphorylation Site (EGF→HRG, 24H)	Measurement Type
FAK Y 576	10 minutes
P38 A Y 182	30 minutes
p130Cas Y 249	30 minutes
p130Cas Y 234	30 minutes
IGF1R Y 1161	30 minutes
GRF1 Y 1105	30 minutes
p130Cas Y 249	10 minutes
paxillin Y 118	5 minutes
BCAR3 Y267	30 minutes
Src Y 418	30 minutes
p130Cas Y 387	10 minutes
p130Cas Y 387	30 minutes
FAK Y 576	5 minutes
Src Y 418	5 minutes
FAK Y 576	30 minutes
p130Cas Y 234	5 minutes
p130Cas Y 387	5 minutes
BCAR3 Y267	5 minutes
p130Cas Y 327	5 minutes
p130Cas Y 249	5 minutes

Table 5-3A

Phosphorylation Site (P→24H, HRG)	Measurement Type
p130Cas Y 327	10 minutes
p130Cas Y 327	30 minutes
PRP4K Y 849	30 minutes
PTPRA Y 798	30 minutes
p130Cas Y 387	10 minutes
PTPRA Y 798	5 minutes
p130Cas Y 234	10 minutes
PI3KR2 Y 464	30 minutes
PTPRA Y 798	10 minutes
SHP-2 Y 62	5 minutes
EphA2 Y/Y 588/594	5 minutes
SHP-2 Y 62	10 minutes
p130Cas Y 387	30 minutes
FAK Y 397	30 minutes
PRP4K Y 849	5 minutes
EphA2 Y 575	10 minutes
EphA2 Y 588	30 minutes
SHP-2 Y 62	30 minutes
RAIG1 Y 347	5 minutes
HER2 Y 1248	30 minutes

Table 5-3B

Phosphorylation Site (24H→P, HRG)	Measurement Type
KIAA1217 Y 393	10 minutes
IGF1R Y 1161	30 minutes
paxillin S/Y 84/88	30 minutes
Dsc3a Y 818	integral
Caveolin 1 Y 14	10 minutes
CrkL Y 132	10 minutes
EGFR Y 1148	10 minutes
GIT1 Y 545	5 minutes
STAT3-2 Y 704	30 minutes
GRF1 Y 1105	10 minutes
IGF1R Y 1165	30 minutes
EGFR Y 1173	30 minutes
Src Y 418	10 minutes
EGFR Y 1173	5 minutes
EGFR Y 1173	integral
EGFR Y 1173	10 minutes
Src Y 418	5 minutes
Dsc3a Y 818	30 minutes
Dsc3a Y 818	10 minutes
Src Y 418	30 minutes

Table 5-4A

Phosphorylation Site (migration-specific)	Measurement Type
GRF1 Y 1105	5 minutes
LDLR Y 845	10 minutes
EphA2 Y/Y 588/594	30 minutes
Erbin Y 1104	5 minutes
An A2 Y 23	5 minutes
Erbin Y 1104	10 minutes
An A2 Y 237	10 minutes
EphA2 Y 588	10 minutes
PZR Y 263	5 minutes
SHB Y 355	5 minutes
Src Y 418	5 minutes
TfR Y 20	10 minutes
An A2 Y 237	5 minutes
HER2 Y 1248	integral
HER2 Y 1248	5 minutes
PI3KR2 Y 464	5 minutes
ITGB4 Y 1207	10 minutes
An A2 Y 237	30 minutes
PTPRF Y 308	5 minutes
KIAA1217 Y 393	5 minutes

Table 5-4B

Phosphorylation Site (proliferation-specific)	Measurement Type
Dsc3a Y 818	30 minutes
Dsc3a Y 818	10 minutes
Dsc3a Y 818	integral
EGFR Y 1173	integral
EGFR Y 1173	10 minutes
paxillin S/Y 84/88	30 minutes
EGFR Y 1173	30 minutes
paxillin S/Y 84/88	integral
Catenin d1 Y 228	integral
CrkL Y 132	integral
Catenin d1 Y 228	30 minutes
CrkL Y 132	10 minutes
paxillin S/Y 84/88	5 minutes
EGFR Y 1173	5 minutes
GRF1 Y 1105	integral
p130Cas Y 327	10 minutes
P38 A Y 182	5 minutes
EphB1 Y 600	integral
p130Cas Y 387	10 minutes
P38 A Y 182	integral

Table 5-4C

Phosphorylation Site (migration & proliferation)	Measurement Type
p130Cas Y 327	10 minutes
IGF1R Y 1165	30 minutes
Src Y 418	30 minutes
GIT1 Y 545	5 minutes
IGF1R Y 1161	30 minutes
SHP-2 Y 62	30 minutes
FAK Y 576	5 minutes
EphA2 Y 575	5 minutes
EphA2 Y/Y 588/594	5 minutes
SHP-2 Y 62	10 minutes
p130Cas Y 327	30 minutes
SHP-2 Y 62	5 minutes
PRP4K Y 849	30 minutes
Src Y 418	integral
HER2 Y 1248	30 minutes
Dsc3a Y 818	10 minutes
An A2 Y 237	30 minutes
Dsc3a Y 818	30 minutes
EphA2 Y 575	10 minutes
EGFR Y 1173	5 minutes

Table 5-4D

Phosphorylation Site (least important)	Measurement Type
FAK Y 397	5 minutes
CDK2 Y 15	5 minutes
EphA2 Y 772	30 minutes
BCAR3 Y267	10 minutes
paxillin Y 118	30 minutes
p130Cas Y 234	30 minutes
FAK Y 397	30 minutes
RAIG1 Y 347	10 minutes
ERK2 Y 187	10 minutes
IGF1R Y 1161	5 minutes
SH2D3A S/Y 218/231	10 minutes
BCAR3 Y267	30 minutes
ERK1 Y 204	30 minutes
GIT1 Y 545	10 minutes
paxillin Y 118	10 minutes
PTPRA Y 798	integral
SH2D3A S/Y 218/231	5 minutes
RAIG1 Y 347	5 minutes
ERK1 Y 204	10 minutes
p130Cas Y 327	integral

References

1. Mumby, M. and D. Brekken, *Phosphoproteomics: new insights into cellular signaling.* Genome Biol, 2005. **6**(9): p. 230.
2. Schmelzle, K. and F.M. White, *Phosphoproteomic approaches to elucidate cellular signaling networks.* Curr Opin Biotechnol, 2006. **17**(4): p. 406-14.
3. Gruhler, A., et al., *Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway.* Mol Cell Proteomics, 2005. **4**(3): p. 310-27.
4. Kratchmarova, I., et al., *Mechanism of divergent growth factor effects in mesenchymal stem cell differentiation.* Science, 2005. **308**(5727): p. 1472-7.
5. Hynes, N.E. and D.F. Stern, *The biology of erbB-2/neu/HER-2 and its role in cancer.* Biochim Biophys Acta, 1994. **1198**(2-3): p. 165-84.
6. Citri, A. and Y. Yarden, *EGF-ERBB signalling: towards the systems level.* Nat Rev Mol Cell Biol, 2006. **7**(7): p. 505-16.
7. Yarden, Y. and M.X. Sliwkowski, *Untangling the ErbB signalling network.* Nat Rev Mol Cell Biol, 2001. **2**(2): p. 127-37.
8. Neve, R.M., T. Holbro, and N.E. Hynes, *Distinct roles for phosphoinositide 3-kinase, mitogen-activated protein kinase and p38 MAPK in mediating cell cycle progression of breast cancer cells.* Oncogene, 2002. **21**(29): p. 4567-76.
9. Spencer, K.S., et al., *ErbB2 is necessary for induction of carcinoma cell invasion by ErbB family receptor tyrosine kinases.* J Cell Biol, 2000. **148**(2): p. 385-97.
10. Wolf-Yadlin, A., et al., *HER2-overexpression effects on cell signaling networks governing proliferation and migration.* Mol Syst Biol, 2006. **in press**.

11. Geladi, P. and B. Kowalski, *Partial Least Squares Regression: A Tutorial.* Analytica Chimica Acta, 1986. **185**: p. 1-17.

12. Hoskuldsson, A., *PLS Regression Methods.* Journal of Chemometrics, 1988. **2**: p. 211-228.

13. Eriksson, L., et al., *Multi- and Megavariate Data Analysis: Principles and Applications.* 2001, Umea, Sweden: Umetrics.

14. Zamir, E. and B. Geiger, *Components of cell-matrix adhesions.* J Cell Sci, 2001. **114**(Pt 20): p. 3577-9.

15. Kumar, N., et al., *Modeling HER2 effects on cell behavior from mass spectrometry phosphotyrosine data.* PLoS Computational Biology, 2006. **in press**.

16. Rothhut, B., *Participation of annexins in protein phosphorylation.* Cell Mol Life Sci, 1997. **53**(6): p. 522-6.

17. Liu, J.W., et al., *Annexin II expression is reduced or lost in prostate cancer cells and its re-expression inhibits prostate cancer cell migration.* Oncogene, 2003. **22**(10): p. 1475-85.

18. Tatenhorst, L., et al., *Knockdown of annexin 2 decreases migration of human glioma cells in vitro.* Neuropathol Appl Neurobiol, 2006. **32**(3): p. 271-7.

19. Wang, F.M., et al., *SHP-2 promoting migration and metastasis of MCF-7 with loss of E-cadherin, dephosphorylation of FAK and secretion of MMP-9 induced by IL-1beta in vivo and in vitro.* Breast Cancer Res Treat, 2005. **89**(1): p. 5-14.

20. Burkart, A., et al., *Regulation of the SHP-2 tyrosine phosphatase by a novel cholesterol- and cell confluence-dependent mechanism.* J Biol Chem, 2003. **278**(20): p. 18360-7.

21. Wright, J.D., C.W. Reuter, and M.J. Weber, *Identification of sites on epidermal growth factor receptors which are phosphorylated by pp60src in vitro.* Biochim Biophys Acta, 1996. **1312**(2): p. 85-93.

22. Keilhack, H., et al., *Phosphotyrosine 1173 mediates binding of the protein-tyrosine phosphatase SHP-1 to the epidermal growth factor receptor and attenuation of receptor signaling.* J Biol Chem, 1998. **273**(38): p. 24839-46.

23. Defilippi, P., P. Di Stefano, and S. Cabodi, *p130Cas: a versatile scaffold in signaling networks.* Trends Cell Biol, 2006. **16**(5): p. 257-63.

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

25. Ravichandran, K.S., *Signaling via Shc family adapter proteins.* Oncogene, 2001. **20**(44): p. 6322-30.

26. Reeves, S.A., et al., *Developmental regulation of annexin II (Lipocortin 2) in human brain and expression in high grade glioma.* Cancer Res, 1992. **52**(24): p. 6871-6.

27. Jing, S.Q., et al., *Role of the human transferrin receptor cytoplasmic domain in endocytosis: localization of a specific signal sequence for internalization.* J Cell Biol, 1990. **110**(2): p. 283-94.

28. Carlevaro, M.F., et al., *Transferrin promotes endothelial cell migration and invasion: implication in cartilage neovascularization.* J Cell Biol, 1997. **136**(6): p. 1375-84.

29. O'Donnell, K.A., et al., *Activation of transferrin receptor 1 by c-Myc enhances cellular proliferation and tumorigenesis.* Mol Cell Biol, 2006. **26**(6): p. 2373-86.

30. Johannessen, L.E., et al., *Activation of the epidermal growth factor (EGF) receptor induces formation of EGF receptor- and Grb2-containing clathrin-coated pits.* Mol Cell Biol, 2006. **26**(2): p. 389-401.

31. Puri, C., et al., *Relationships between EGFR signaling-competent and endocytosis-competent membrane microdomains.* Mol Biol Cell, 2005. **16**(6): p. 2704-18.

32. Warren, R.A., F.A. Green, and C.A. Enns, *Saturation of the endocytic pathway for the transferrin receptor does not affect the endocytosis of the epidermal growth factor receptor.* J Biol Chem, 1997. **272**(4): p. 2116-21.

33. Yang, W., et al., *The Cdc42 target ACK2 directly interacts with clathrin and influences clathrin assembly.* J Biol Chem, 2001. **276**(20): p. 17468-73.

34. Coon, M. and R. Herrera, *Modulation of HeLa cells spreading by the non-receptor tyrosine kinase ACK-2.* J Cell Biochem, 2002. **84**(4): p. 655-65.

35. Galisteo, M.L., et al., *Activation of the nonreceptor protein tyrosine kinase Ack by multiple extracellular stimuli.* Proc Natl Acad Sci U S A, 2006. **103**(26): p. 9796-801.

36. Dyson, J.M., et al., *The SH2-containing inositol polyphosphate 5-phosphatase, SHIP-2, binds filamin and regulates submembraneous actin.* J Cell Biol, 2001. **155**(6): p. 1065-79.

37. Habib, T., et al., *Growth factors and insulin stimulate tyrosine phosphorylation of the 51C/SHIP2 protein.* J Biol Chem, 1998. **273**(29): p. 18605-9.

38. Prasad, N.K. and S.J. Decker, *SH2-containing 5'-inositol phosphatase, SHIP2, regulates cytoskeleton organization and ligand-dependent down-regulation of the epidermal growth factor receptor.* J Biol Chem, 2005. **280**(13): p. 13129-36.

39. Bode, B.P., *Recent molecular advances in mammalian glutamine transport.* J Nutr, 2001. **131**(9 Suppl): p. 2475S-85S; discussion 2486S-7S.

40. Hyde, R., K. Peyrollier, and H.S. Hundal, *Insulin promotes the cell surface recruitment of the SAT2/ATA2 system A amino acid transporter from an endosomal compartment in skeletal muscle cells.* J Biol Chem, 2002. **277**(16): p. 13628-34.

41. Cabodi, S., et al., *p130Cas as a new regulator of mammary epithelial cell proliferation, survival, and HER2-neu oncogene-dependent breast tumorigenesis.* Cancer Res, 2006. **66**(9): p. 4672-80.

42. Cox, B.D., et al., *New concepts regarding focal adhesion kinase promotion of cell migration and proliferation.* J Cell Biochem, 2006.

43. Mitra, S.K., et al., *Intrinsic focal adhesion kinase activity controls orthotopic breast carcinoma metastasis via the regulation of urokinase plasminogen activator expression in a syngeneic tumor model.* Oncogene, 2006. **25**(32): p. 4429-40.

44. Planas-Silva, M.D., et al., *Role of c-Src and focal adhesion kinase in progression and metastasis of estrogen receptor-positive breast cancer.* Biochem Biophys Res Commun, 2006. **341**(1): p. 73-81.

45. Pelkmans, L., et al., *Genome-wide analysis of human kinases in clathrin- and caveolae/raft-mediated endocytosis.* Nature, 2005. **436**(7047): p. 78-86.

46. Polo, S. and P.P. Di Fiore, *Endocytosis conducts the cell signaling orchestra.* Cell, 2006. **124**(5): p. 897-900.

47. Cully, M., et al., *Beyond PTEN mutations: the PI3K pathway as an integrator of multiple inputs during tumorigenesis.* Nat Rev Cancer, 2006. **6**(3): p. 184-92.

48. Oda, K., et al., *A comprehensive pathway map of epidermal growth factor receptor signaling.* Mol Syst Biol, 2005. **1**: p. 2005 0010.

49. Martys, J.L., et al., *Wortmannin-sensitive trafficking pathways in Chinese hamster ovary cells. Differential effects on endocytosis and lysosomal sorting.* J Biol Chem, 1996. **271**(18): p. 10953-62.

Chapter 6 The importance of signal state off-target effects for the pharmacological intervention of HER2-mediated migration

In this chapter, we use systems principles (Chapters 4 and 5) to further our understanding of drug efficacy in HER2 overexpressing cells.

6.1 Introduction

The effect of increased signal transduction through a particular intracellular kinase on cell function is dependent on the network signal state within which the change occurs [1, 2]. By extension, the effect of attenuating kinase activity on cell function by treatment with a small molecule kinase inhibitor is also dependent on the signal state. Further complicating the matter is the fact that treatment with inhibitor or drug is likely to change not only the kinase activity of the target but also many elements of the greater signal state. This effect, often referred to as an 'off-target' drug effect, can be due to the promiscuous binding of the drug or the interconnectedness of the signaling network. A great deal of effort has been spent identifying these off-target effects, with major advances coming in the identification of both off-target binding sites and changes in signal state due to intracellular crosstalk [2, 3]. How these off-target events contribute to a drug's control of cell function, however, is not well understood. In this study, we explored the effects of changes in signal state due to treatment with three small molecule inhibitors, LY294002, PD98059, and Gefitinib (Iressa), on human epidermal

growth factor receptor 2 (HER2)-mediated cell migration in human mammary epithelial cells (HMEC).

HER2 is overexpressed in 20% to 30% of breast cancers and correlates with increased metastasis and poor prognosis [4]. It is a member of the ErbB or HER receptor family (comprised of HER1/Epidermal Growth Factor Receptor (EGFR), HER2, HER3, and HER4), and its activation is coordinated through concentration-dependent homodimerization or ligand-driven heterodimerization with other HER family receptors. Epidermal growth factor (EGF) and heregulin (HRG), two ErbB family ligands implicated in cancer progression, bind HER1 and HER3, respectively, to induce the activation of HER2 through heterodimerization [5].

Because of HER2's role in breast cancer metastasis, a number of groups have investigated the effect of HER2 expression on cell motility, demonstrating increased invasion and motility in HER2 overexpressing breast cancer cell lines [6, 7]. In particular, we have shown that HER2 overexpression increases cell migration in human mammary epithelial cells (HMEC) presented with no ligand, EGF, or HRG, by increasing the directional persistence of cell movement, while only increasing speed under EGF treatment [6]. Identification of persistence as the critical mechanism driving increased migration is consistent with prior studies of primary ductal breast carcinoma cells and neuroepithelial tumors revealing elevated directional persistence as a hallmark of cancer cell movement [8, 9]. Despite its importance in cancer cell migration, the network of signals that give rise to directional persistence and cell speed, especially in the absence of extracellular gradients, are not well characterized. We sought to characterize elements of this network by quantifying the phosphorylation of Akt (serine

473), Erk (threonine 202/tyrosine 204), EGFR (tyrosine 1173 and tyrosine 1068), and p38 (threonine 180/tyrosine 182) in parallel with measurements of cell speed and persistence in HMEC cells under varying ligand and inhibitor treatments. The kinases were chosen for three reasons: 1. they have all been shown to regulate cell migration in breast cancer epithelial cells [7, 10-12], 2. they have all been implicated in the control of directional migration [13-16], and 3. the downstream kinases are highly informative integrators of intracellular signals as suggested by a recent mapping of the EGFR signaling network [17]. We chose three small molecule inhibitors, each of which targeted a measured kinase signal and had a well-defined binding profile [3, 18]. PD98059 is a MEK inhibitor that reduces ERK activity with high specificity, LY294002 inhibits phosphatidylinositol 3-kinase (PI3K) which leads to decreases in Akt phosphorylation, and Gefitinib is a clinically approved EGFR inhibitor known to suppress EGFR phosphorylation.

In this paper, we show that the measurement of highly informative kinase signaling nodes, representing both on- and off-target molecules for the inhibitors studied, is critical for an understanding of drug efficacy against cell speed and persistence as HER2 expression is varied. Migration measurements revealed a ligand-dependent efficacy of inhibitors against directional persistence, and demonstrated that Gefitinib is a potent inhibitor of cell migration and directional persistence in HER2 overexpressing systems. Signal profiling in response to drug treatment revealed off-target effects such as the inhibition of Erk phosphorylation in response to PI3K inhibition and the decrease of Y1173 EGFR phosphorylation in response to MEK inhibition. A computational linear model helped codify the relationship between signal state and cell

migration and revealed the importance of multiple kinase measurements for the accurate prediction of the effect of drugs on migration.

6.2 Materials and Methods

6.2.1 Cell culture and stimulation

184A1 human mammary epithelial cells were used as described previously [6, 19]. The HMEC parental cells were a kind gift from Martha Stampfer (Lawrence Berkeley Laboratory, Berkeley, CA) and were cultured in DFCI-1 media supplemented with 12.5 ng/ml EGF (Peprotech, Inc.) [20]. The 184A1 HMEC clone 24H cells were a kind gift from Steve Wiley (Pacific Northwest National Laboratories, Richland WA) and were maintained in DFCI-1 media supplemented with 12.5 ng/ml EGF and 150 μg/ml Geneticin (Gibco, Invitrogen Inc.). We defined our serum free DFCI-1 media as stated in [6], without EGF, bovine pituitary extract, fetal bovine serum, and insulin. Note that this serum-free formulation is different than that in [19] due to addition of insulin in serum-free media in [19]. Receptor numbers and dimer populations in response to EGF and HRG stimulation are given in [19]. For signaling and migration assays, cells were stimulated with either 100 ng/ml EGF or 80 ng/ml HRG-β1 (Sigma, Inc.). For inhibitor experiments, cells were pre-treated with either 20 μM LY294002 (Calbiochem, Inc.) for one hour, 25μM PD98059 (Calbiochem, Inc.) for one hour, 20 μM Gefitinib (WuXi Pharmatech Co., Ltd) for 30 minutes, or 200 nM Wortmannin (Calbiochem, Inc.) for 20 minutes.

6.2.2 Immunocytochemistry

Cells were seeded at ~50,000 cells/well in a 96-well plate (NUNC 165305) in full DFCI-1 media. After 4-6 hours full media was replaced with serum-free media. After 12-16 hours of serum starvation, cells were pre-incubated with inhibitor or treated with EGF, HRG, or fresh serum free media. A wound of width ~650 μm was then scraped in each well. Cells were stimulated for 0 (serum free), 5, 15, 30, 60, and 90 minutes, with three or four wells used as biological replicates for each condition. After desired time had elapsed, cells were fixed using 3.7% formaldehyde diluted in phosphate buffered saline solution (PBS) for 20 minutes at room temperature. The fix solution was then removed and cells were permeabilized with 0.1% Triton. Cells were then blocked using Odyssey Blocking Buffer (LI-COR Biosciences, Inc.) for 1 hour. Primary anitibody diluted 1:100 in blocking buffer was added to each well and the plate was then place at 4°C overnight. The phospho-p44/42 MAPK (Erk, Thr202/Tyr204, #4377) and phospho-Akt (Ser473, #4058) rabbit monoclonal antibodies were purchased from Cell Signaling, Inc. The phospho-p38 MAPK (pT180/pY182, #1229), phospho-EGFR (pY1068, #1138), and EGFR (pY1173, #1124) rabbit monoclonal antibodies were purchased from Epitomics, Inc. After overnight incubation, the plate was washed with 0.1% Tween-20 in PBS, and then incubated at room temperature for one hour with secondary antibody solution [goat anti-rabbit IRDye 800 (LI-COR Biosciences, Inc.) diluted 1:800 in blocking buffer] along with the DNA stain TO-PRO-3 (diluted 1:2500 in blocking buffer, Invitrogen, Inc.). Plates were then washed again with 0.1% Tween-20 in PBS and immediately imaged at 700nm and 800nm using an Odyssey instrument. The integrated intensity of the IRDye bound to primary antibody (800 nm channel) was normalized to the total DNA integrated

intensity (700 nm channel) for each well. After averaging biological replicate data, 24H and parental data were normalized to the parental serum free value run on the same plate. Typically, only one ligand was used per plate, thus giving rise to two separate normalizations for 24H data, one for the +EGF data and another for the +HRG data. All signaling data are plotted ±SEM, with N = 3 or 4.

All five monoclonal antibodies were validated using western blot analysis for specificity using lysates from parental and 24H cell lines after stimulation with full serum (Figure 6-1). The western blot was performed as described in [21]. Electrophoresis was performed in a 7% polyacrylamide gel and the above indicated primary and secondary antibodies were used for the membrane probe and visualization steps.

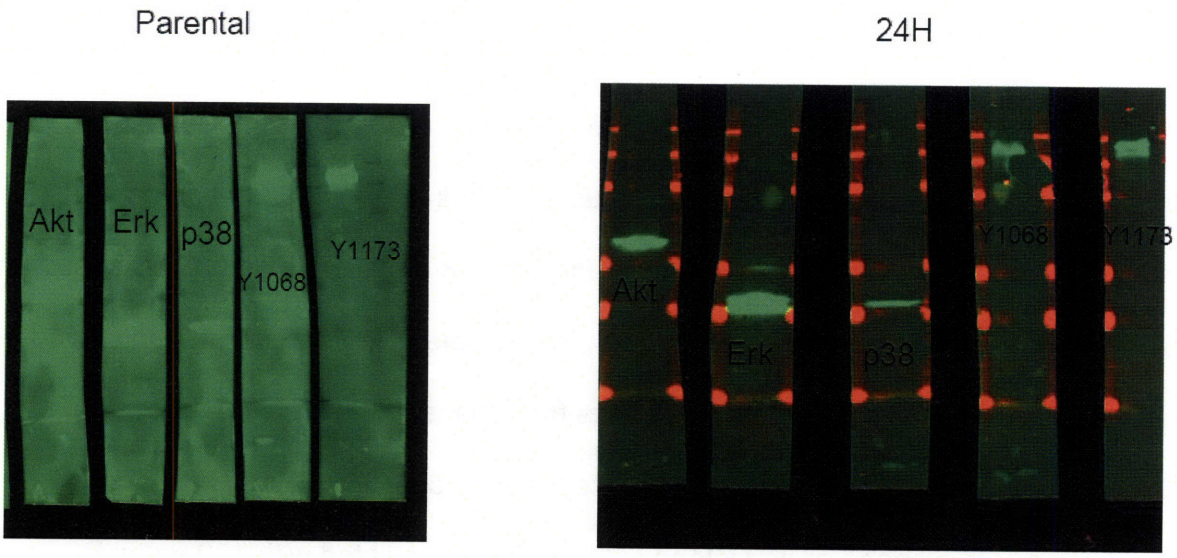


Figure 6-1: Western blotting reveals specificity for monoclonal antibodies against Erk, Akt, p38, EGFR Y1068, and EGFR Y1173. Western blots were performed as described in Materials and Methods to show antibody specificity.

Since this was new signaling platform, we desired to ascertain the consistency of measurement from plate to plate. To do this, we compared normalized serum free

values for 24H cells run on different plates but on the same day. As can be seen in Figure 6-7, we observed a very tight correlation between the 'two' serum free conditions for 24H, i.e. the +HRG 24H and the +EGF 24H conditions. Ideally, the two measurements would be the same after normalization to total DNA and then to parental serum free levels. In practice, the variation was quite low, typically less than 10% of the reported serum free value, except in the case of EGFR Y1173 under no inhibitor.

6.2.3 Migration assay

The migration assay was performed as previously described [6]. Briefly, cells labeled with 5-chloromethylfluorescein diacetate (CMFDA, Molecular Probes, Inc.) were diluted 1:20 with unlabelled cells and plated as described above, treated with ligand and inhibitor as described above, wounded, and cell movement was tracked for 12 hours at 37°C and 5% CO_2 using an automated epiflourescent microscope. Individual cell trajectories in a monolayer were analyzed using Imaris (Bitplane, Inc.) and cell trajectories were then fit to the persistent random walk equation to derive cell speed and persistence. For each condition measured, four separate wells were run and served as four biological replicates. Individual cell data from these four wells were pooled to calculate an average speed and persistence for each condition. All averages were normalized by the speed or persistence of 24H cells stimulated with 100 ng/ml EGF on the plate, to account for plate to plate variability. For each condition, $N > 300$ and all data are reported as ± SEM.

6.2.4 Linear modeling using partial least squares regression (PLSR)

To solve the equation below, where **X** is rank deficient, we implemented partial least squares regression (PLSR) using SIMCA-P 11.0 (Umetrics) as previously described [19].

$$\mathbf{P} = \mathbf{X}c$$

As described in the Results section, **P** is a column vector (*Mx1*) of persistence values corresponding to *M* cellular conditions, **X** is the signaling matrix (*MxN*) comprised of *M* row vectors, for which each column represents a signaling metric (i.e. Akt phosphorylation at 5 minutes), and *c* is a vector of constants (*Nx1*) that maps **X** onto **P**. The metrics included for each phosphorylation site were the normalized levels of phosphorylation at each time point as well as the integral of phosphorylation to serve as a metric for 'net' phosphorylation over 90 minutes. Coefficients for each metric were extracted from SIMCA-P and correspond to the coefficients in the model when X is scaled and centered and P is scaled. They are calculated as described elsewhere [22]. All data was mean-centered and scaled to unit variance prior to analysis.

6.3 Results

6.3.1 Investigations into the role of Akt phosphorylation on cell persistence reveal the need for signal state measurements

Akt and its upstream kinase PI3K have been identified as critical for directionally persistent motion in chemotaxing cells [16]. We hypothesized that directional movement in the absence of chemical gradients might be coordinated by signaling features common to chemotaxing cells. Thus, we studied the role of Akt signaling and inhibition in HER2-mediated increases in directional movement.

As previously demonstrated, HMEC cells overexpressing HER2 (24H) had higher levels of directional persistence than those with low HER2 expression (Parental cells) in the presence or absence of EGF or HRG (Figure 6-2A). Investigation of Akt phosphorylation under these same conditions revealed increased phosphorylation in response to HER2 overexpression, suggesting a potential role for Akt in persistent movement (Figure 6-2B). Inhibition of Akt by treatment with LY294002 decreased persistence in 24H cells treated with HRG, further corroborating the role of the PI3K/Akt pathway in directional persistence (Figure 6-2C). Inhibition with LY294002 in 24H cells treated with EGF, however, had no effect on directional persistence (Figure 6-2D). Quantification of Akt phosphorylation in response to inhibition revealed consistent knockdown for both HRG and EGF treatments (Figure 6-2E). Thus, inhibition of persistence was ligand-dependent, suggesting that the effect of Akt inhibition on persistence depends on the signaling context within which the inhibition takes place.

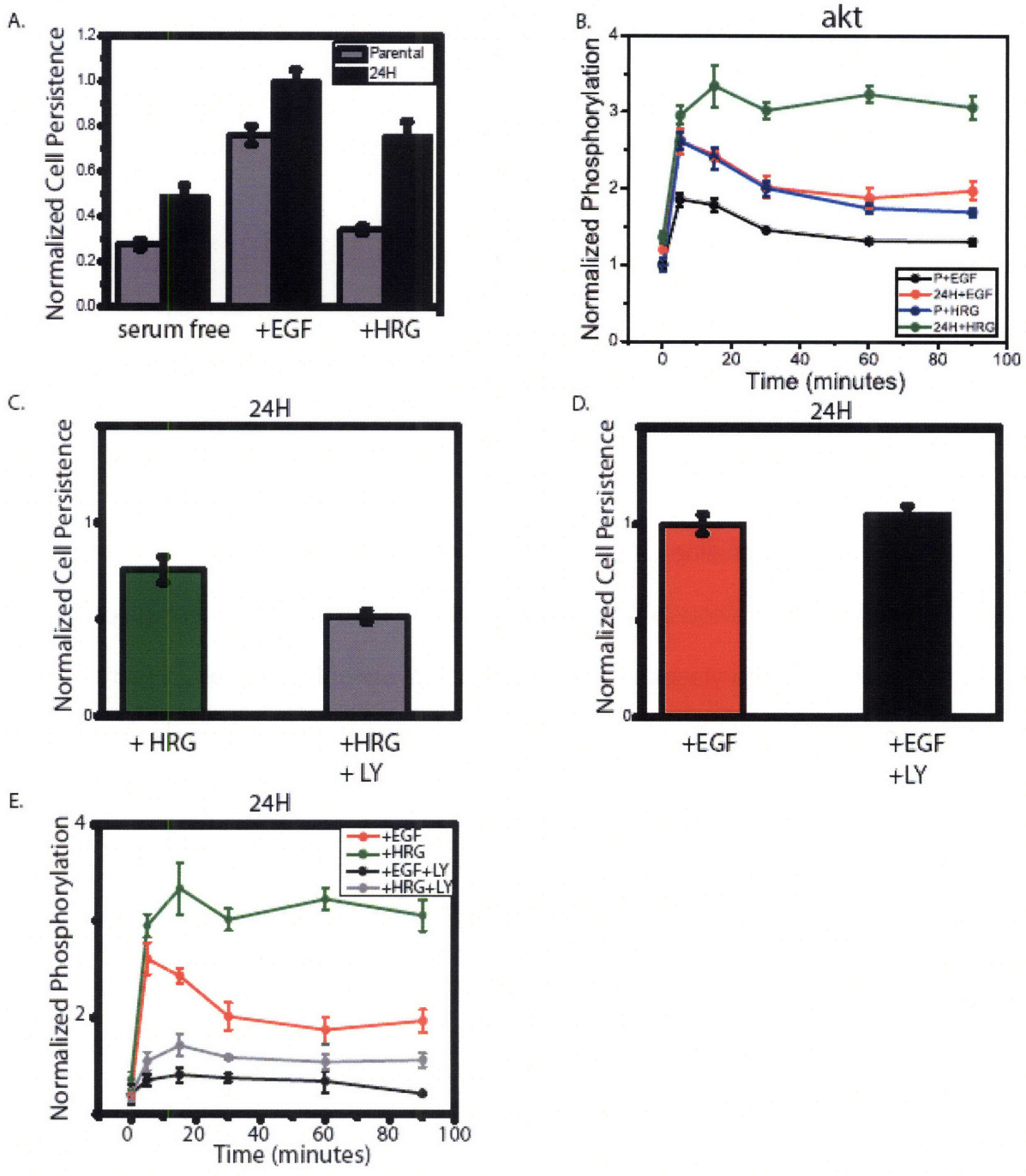


Figure 6-2: Akt inhibition by LY294002 decreases persistence in HRG stimulated 24H cells but not in EGF stimulated 24H cells. (A) Directional persistence was previously determined and replotted here normalized to the 24H+EGF value (ref) for parental cells (gray) and 24H cells (black) in the presence of EGF (100 ng/ml), HRG (80 ng/ml), or serum-free conditions. (B) A immunocytochemical technique quantified pAkt (S473) phosphorylation in parental and 24H cells stimulated with EGF (100 ng/ml) or HRG (80 ng/ml). (C) Directional persistence of 24H cells stimulated with HRG (80

ng/ml) in the presence or absence of pretreatment with LY294002 (20μM). (D) Directional persistence of 24H cells stimulated with EGF (100 ng/ml) in the presence or absence of pretreatment with LY294002 (20μM). (E) Phosphorylation of Akt (S473) quantified in 24H cells stimulated with EGF (100 ng/ml) or HRG (80 ng/ml) in the presence or absence of pretreatment with LY294002 (20μM). All data shown ±SEM and normalized as indicated in Materials and Methods.

6.3.2 A high-throughput immunocytochemical technique quantifies Erk, Akt, EGFR, and p38 phosphorylation in a scratch assay format

To quantify cell signaling under the same conditions studied in the migration assay, we employed a previously published high-throughput immunocytochemical assay to quantify phosphorylation in wounded monolayers of HMEC cells (Figure 6-3A and Materials and Methods). We selected three kinases, Erk (T202/Y204), Akt (S473), and p38 (T180/Y182), that have a demonstrated ability to control cell migration and also serve as integrators of signaling information in EGFR signaling networks [17]. In addition, we profiled phosphorylation at two EGFR sites, Y1068 and Y1173, both known to stimulate multiple downstream signaling pathways, but whose individual function, especially in the control of cell migration, is not well understood [23]. HMEC cells were seeded in a 96-well plate, wounded in the presence or absence of HRG (80 ng/ml) or EGF (100 ng/ml), and then assayed for kinase phosphorylation at 0, 5, 15, 30, 60, and 90 minutes (Figure 6-3).

EGF treatment resulted in the phosphorylation of Erk in both Parental and 24H cells, with higher initial and sustained activation occurring in the 24H cells. In contrast, HRG treatment did not stimulate Erk phosphorylation in either cell line, indicating that

the low levels of HER2-HER3 heterodimers formed were not effective in promoting signaling through this pathway (Figure 6-3B, see ref for explanation and calculation of numbers of dimers formed in response to ligand treatment). As mentioned above, treatment with HRG did stimulate Akt, and HER2 overexpression lead to a more sustained response in 24H cells as compared to parental cells (Figure 6-3C). EGF treatment also promoted Akt phosphorylation that increased with HER2 expression, although in the case of EGF treatment, the dynamics of phosphorylation were similar, with both parental and 24H cells exhibiting similar amounts of dephosphorylation over the 90 minute time course (Figure 6-3C). Phosphorylation at EGFR Y1068 and Y1173 rose in response to EGF but not to HRG, as expected. In the case of the Y1068 site, we observed elevated levels of phosphorylation in 24H cells at the zero minute time point (serum-free condition), suggesting the existence of an autocrine loop. 24H cells exhibited increased absolute levels of Y1068 phosphorylation in response to EGF treatment as compared with the parental cells. Both 24H and parental cells, however, had similar phosphorylated levels of Y1173. Finally, p38 phosphorylation was higher in 24H cells but remained relatively constant in response to HRG and EGF, although we observed a slight peak at 15 minutes under all conditions (only the 15 minute HRG-stimulated peaks are significantly greater than the zero minute phosphorylation levels at $p \leq 0.05$, Figure 6-3F).

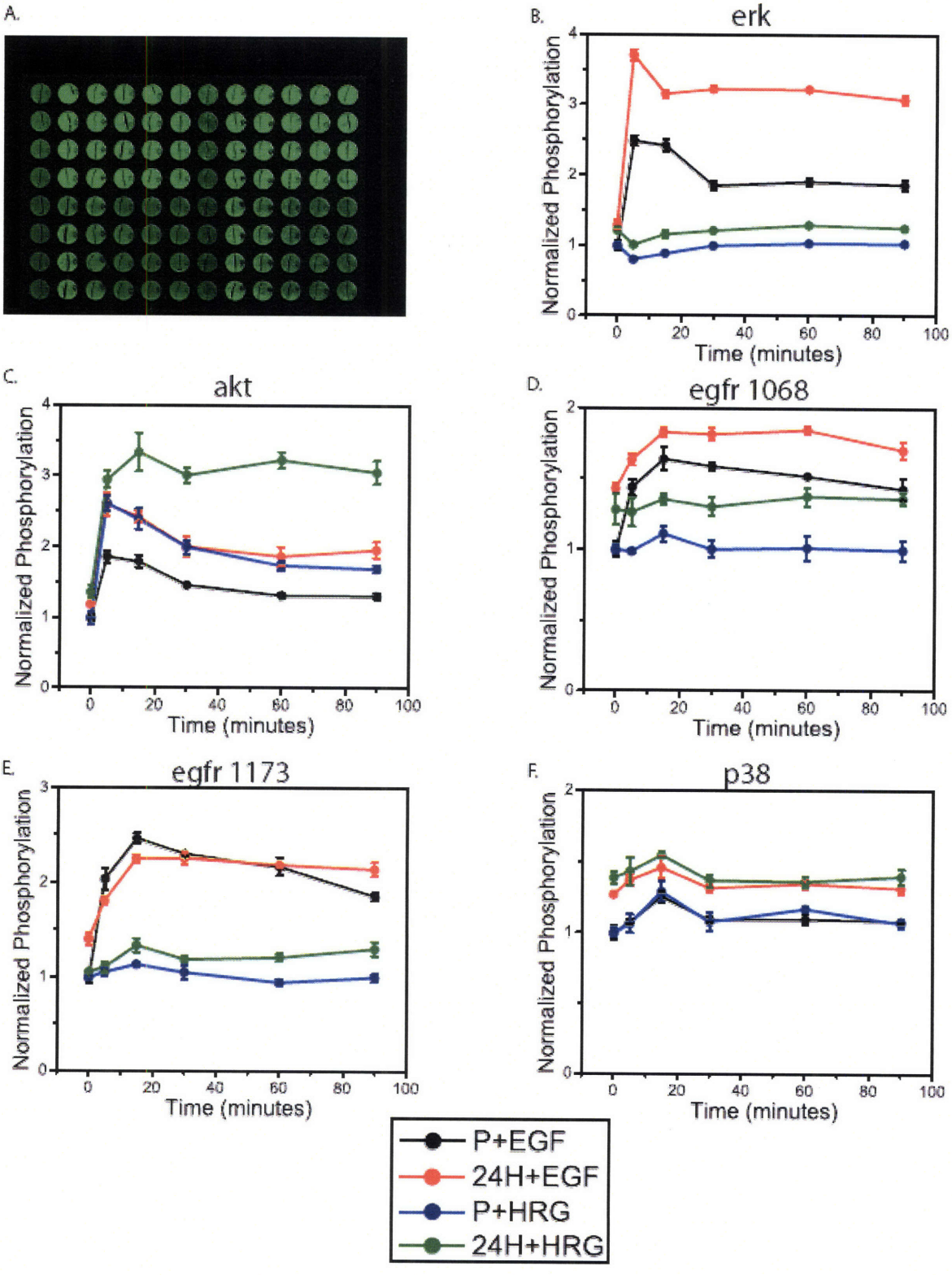

A.
B.
erk
Normalized Phosphorylation
Time (minutes)
C.
akt
Normalized Phosphorylation
Time (minutes)
D.
egfr 1068
Normalized Phosphorylation
Time (minutes)
E.
egfr 1173
Normalized Phosphorylation
Time (minutes)
F.
p38
Normalized Phosphorylation
Time (minutes)
P+EGF
24H+EGF
P+HRG
24H+HRG

Figure 6-3: Quantified phosphorylation for Erk, Akt, p38, and EGFR. (A) An example image of the 96-well immunocytochemical assay performed to quantify phosphorylation. Note the wounded monolayer of cells. Shown here is the 800 nm channel (see Materials and Methods). Quantified phosphorylated Erk (B), Akt (C), EGFR Y1068 (D), EGFR Y1173 (E), and p38 (F) for parental and 24H cells stimulated with EGF (100 ng/ml) or HRG (80 ng/ml). All data is shown ±SEM.

Thus, the quantification of five phosphorylation sites revealed a broad state of kinase activation stimulated by EGF and differentiated by HER2 expression levels. HRG, in contrast, stimulated very little phosphorylation in kinases other than Akt. Consistent increasing levels of serum-free kinase phosphorylation with HER2 expression, particularly apparent in the case of p38, suggest a role for autocrine signaling or HER2 homodimerization in the regulation of cell function.

6.3.3 Migration inhibition by LY294002, PD98059, and Gefitinib exhibit ligand and HER2 dependence

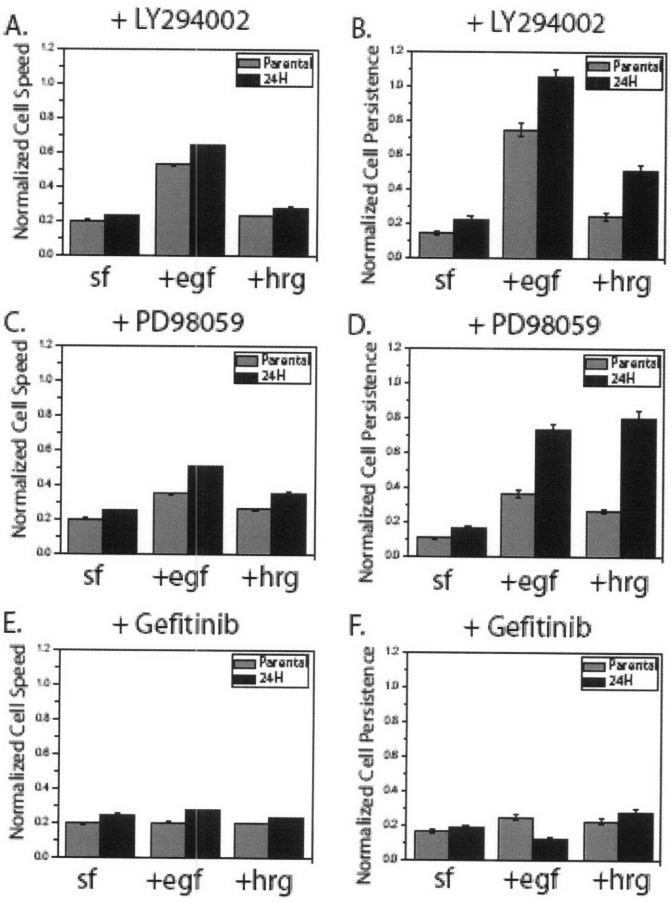
Signaling differences between 24H and parental cells suggested that inhibition of Erk, Akt, and EGFR should differentially inhibit HER2-mediated increases in speed and persistence depending on ligand treatment. We measured cell speed and persistence for each cell line treated with either EGF (100 ng/ml), HRG (80 ng/ml), or no ligand in the presence of either LY294002 (20μM), PD98059 (25 μM), or Gefitinib (20 μM).

Treatment of parental and 24H cells with LY294002 reduced cell speed and persistence under HRG treatment in both cell lines as compared to cells migrating in the absence of inhibitor (Figure 6-4A,B,G). In particular, inhibition decreased persistence levels almost 50% in 24H cells to levels observed in parental cells without inhibitor

(Figure 6-2A, 3G). LY294002 also decreased cell speed in both parental and 24H cells, but with greater efficacy in parental cells, such that 24H cells had higher observed speeds after inhibition, reversing the trend observed in the absence of inhibitor. LY294002 treatment decreased levels of cell speed in EGF-stimulated parental and 24H cells by approximately one-third, but interestingly, did not affect persistence in either cell line (Figure 6-4A,B,G). Inhibition decreased serum-free levels of speed and migration similarly across cell type, reducing both values by approximately one-half.

Inhibition with PD98059, which from signaling assays we expected to affect migration in response to EGF but not HRG, decreased levels of cell speed and persistence in parental cell treated with EGF and HRG. PD98059 reduced persistence in EGF-treated 24H cells with greater efficacy than in parental cells [~ 50% versus ~25% inhibition for Parental and 24H cells, respectively (Figure 6-4D,G)], suggesting that HER2 overexpressing cells are less reliant on ERK for directional persistence. PD98059 treatment of HRG-stimulated 24H cells weakly inhibited speed (Figure 6-4C,G), and did not inhibit persistence (Figure 6-4D,G), further corroborating the finding that directional persistence is more robust to Erk inhibition.

Treatment with Gefitinib lead to the most significant decrease of cell speed and persistence across cell type and ligand treatment (Figure 6-4E,F,G). Inhibition under EGF stimulating conditions lead to ~70-90% reductions in both cell speed and persistence. Interestingly, Gefitinib decreased persistence significantly more in 24H cells (Figure 6-4G). Thus, increasing HER2 expression in the presence of EGF and Gefitinib actually lowered persistence, and 24H cells treated with HRG and Gefitinib had persistence values less than those observed in parental cells absent inhibitor.



G.

Inhibitor	Ligand	% inhibition Speed (P, 24H)	% inhibition Persistence (P, 24H)
LY294002	none	52, 48	49, 54
LY294002	egf	33, 36	2, -1
LY294002	hrg	58, 34	28, 32
PD98059	none	51, 41	61, 65
PD98059	egf	55, 49	51, 26
PD98059	hrg	52, 15	22, -6
Gefitinib	none	55, 45	41, 63
Gefitinib	egf	75, 73	68, 88
Gefitinib	hrg	64, 46	36, 64

Figure 6-4: Effects of LY294002, PD98059, and Gefitinib on cell speed and persistence. Cell speed and directional persistence of parental and 24H cells in the presence of EGF (100 ng/ml), HRG (80 ng/ml), or serum-free media were quantified after pretreatment with one of three inhibitors: LY294002 (20 μM, A, B), PD98059 (25 μM, C, D), and Gefitinib (20μM, E, F). (G) Percent inhibition as compared to the cell speed or persistence value measured without inhibitor for all inhibitor conditions. Speed and persistence data were normalized as described in Materials and Methods and all data is shown ±SEM.

In sum, a study of inhibitor effects on HMEC cell migration revealed that certain inhibitors are effective in the context of only one activating ligand (e.g. LY294002 against persistence), and that certain inhibitors are more effective in the context of low HER2 expression (e.g. PD98059 against persistence) whereas some are more effective in the context of high HER2 expression (e.g. Gefitinib against persistence). Overall, we

identify Gefitinib as a potent inhibitor of persistence across both HRG and EGF stimulation in low and high HER2 expressing conditions.

6.3.4 Quantification of phosphorylation in response to treatment with LY294002, PD98059, and Gefitinib reveals network-wide inhibitor effects

To better understand the above patterns of inhibitor efficacy, we quantified phosphorylation of the four previously measured kinases (Erk, Akt, EGFR, and p38) under the inhibitor conditions explored above. Our results revealed a highly interconnected signaling network, where the action of any given inhibitor had consequences on multiple phosphorylation sites.

Treatment with LY294002 greatly reduced AKT phosphorylation, although small differences between 24H and parental levels remained (Figure 6-7B). Interestingly, inhibitor treatment also eliminated Erk phosphorylation in response to EGF treatment, an unexpected result (Figure 6-7B). To determine whether this effect was due to an off-target binding event specific to LY294002 as opposed to PI3K regulation of Erk phosphorylation, we replicated the experiment using Wortmannin, a different PI3K inhibitor, and observed that it coordinated a similar complete knockdown of Erk phosphorylation (Figure 6-5). Treatment with LY294002 also increased EGF-stimulated phosphorylation at both EGFR Y1068 and Y1173 (Figure 6-7B). This effect was not due to decreased serum-free phosphorylation levels, since serum free levels of phosphorylated EGFR were the same in the presence or absence of inhibitor (Figure 6-6).

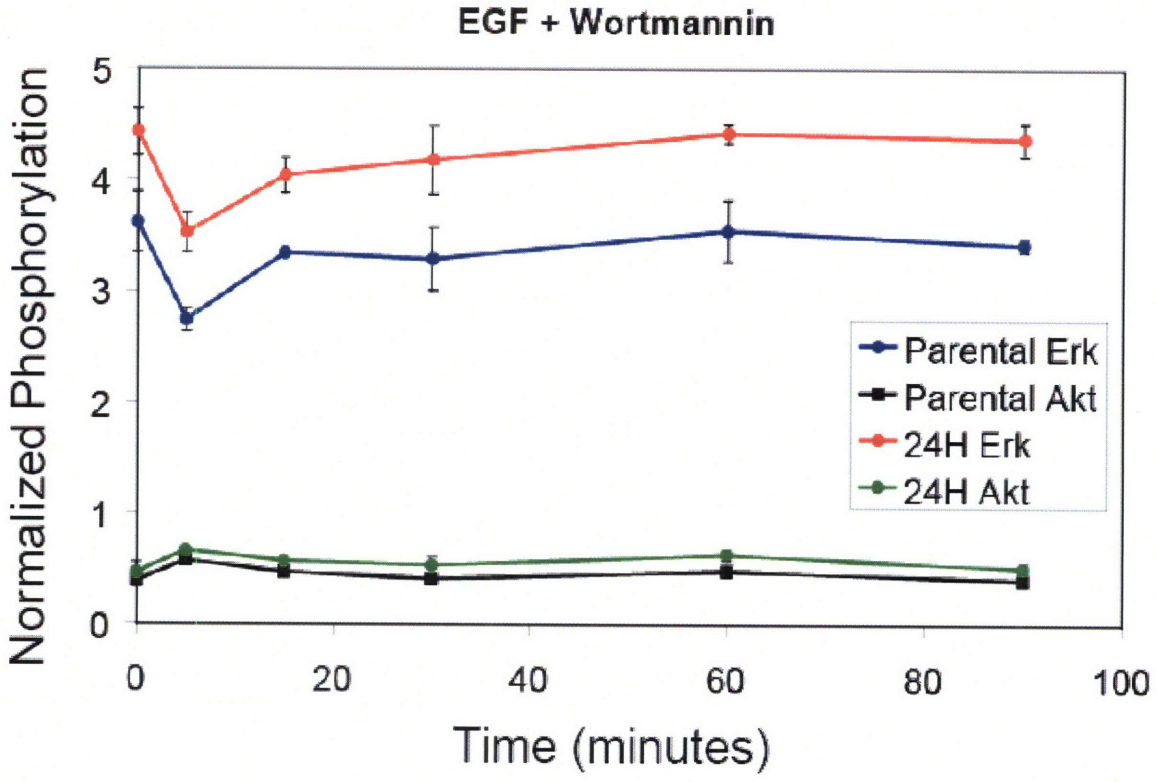


Figure 6-5: Treatment with Wortmannin leads to ablation of Erk signaling. Parental and 24H cells pretreated for 20 minutes with Wortmannin (200 nM) were stimulated with EGF or HRG and the resulting phosphorylation of Erk and Akt was quantified.

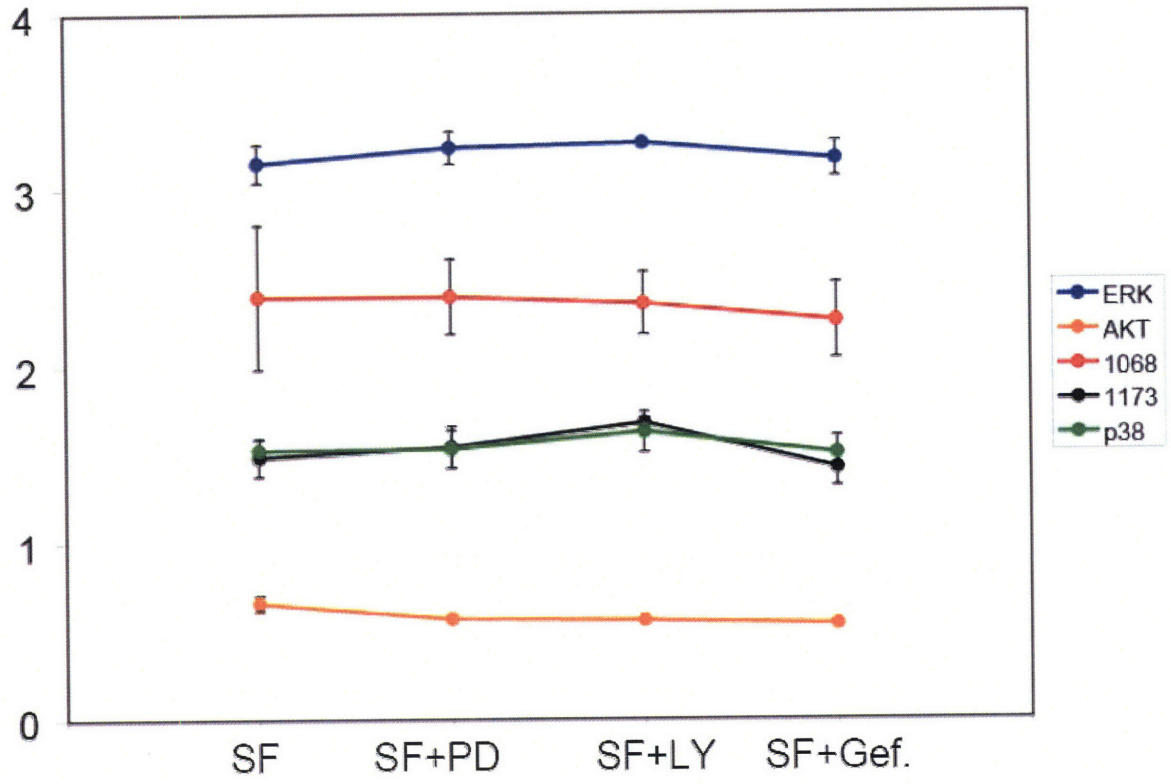


Figure 6-6: Pretreatment with inhibitors does not lead to significant lowering of baseline levels of phosphorylation in Erk, Akt, EGFR, or p38. Cells pretreated with PD98059, LY429002, Gefitinib, or serum-free media, as described in Materials and Methods, were assayed for phosphorylation levels of Erk, Akt, EGFR Y1068, EGFR Y1173, and p38.

Treatment with PD98059 eliminated Erk phosphorylation in response to EGF treatment as expected, and Akt signaling remained relatively consistent with observed phosphorylation in the absence of inhibitor (Figure 6-7C). The difference between 24H and parental p38 phosphorylation levels was decreased over the 90 minute time course due to decreasing 24H phosphorylation concomitant with increasing parental phosphorylation (Figure 6-7C). At 90 minutes p38 phosphorylation in parental cells was higher than that in 24H cells. Phosphorylation at EGFR Y1068 was relatively similar to that observed in cells not treated with inhibitor, although phosphorylation levels in 24H

cells stimulated with EGF were transiently activated, in contrast to the slight but sustained activity observed under normal circumstances (Figure 6-7C). Interestingly, treatment with PD98059 dramatically decreased phosphorylation at EGFR Y1173 in response to EGF. Parental cells exhibited a slight transient spike in response to EGF, but the increase was markedly less than that observed in the absence of inhibitor. 24H cells treated with EGF did not exhibit any measured change in phosphorylation at Y1173, indicating the existence of a potential feedback mechanism between ERK and the Y1173 site.

As expected, Gefitinib treatment eliminated most EGFR phosphorylation in response to EGF (Figure 6-7D). 24H cells treated with EGF, however, did exhibit increased phosphorylation at Y1068 in a manner similar to cells without inhibitor (a small but statistically significant increase in phosphorylation, Figure 6-7D). Parental cells, however, did not respond to EGF stimulation. Erk and Akt phosphorylation were eliminated by Gefitinib treatment, including HRG-stimulated Akt phosphorylation in both parental and 24H cells (Figure 6-7D). In addition, a gradual decrease of p38 phosphorylation levels in 24H cells coupled with a gradual increase in parental cells eliminated the difference observed in p38 phosphorylation levels between the two cell lines in the absence of inhibitor, as observed under PD98059 treatment as well.

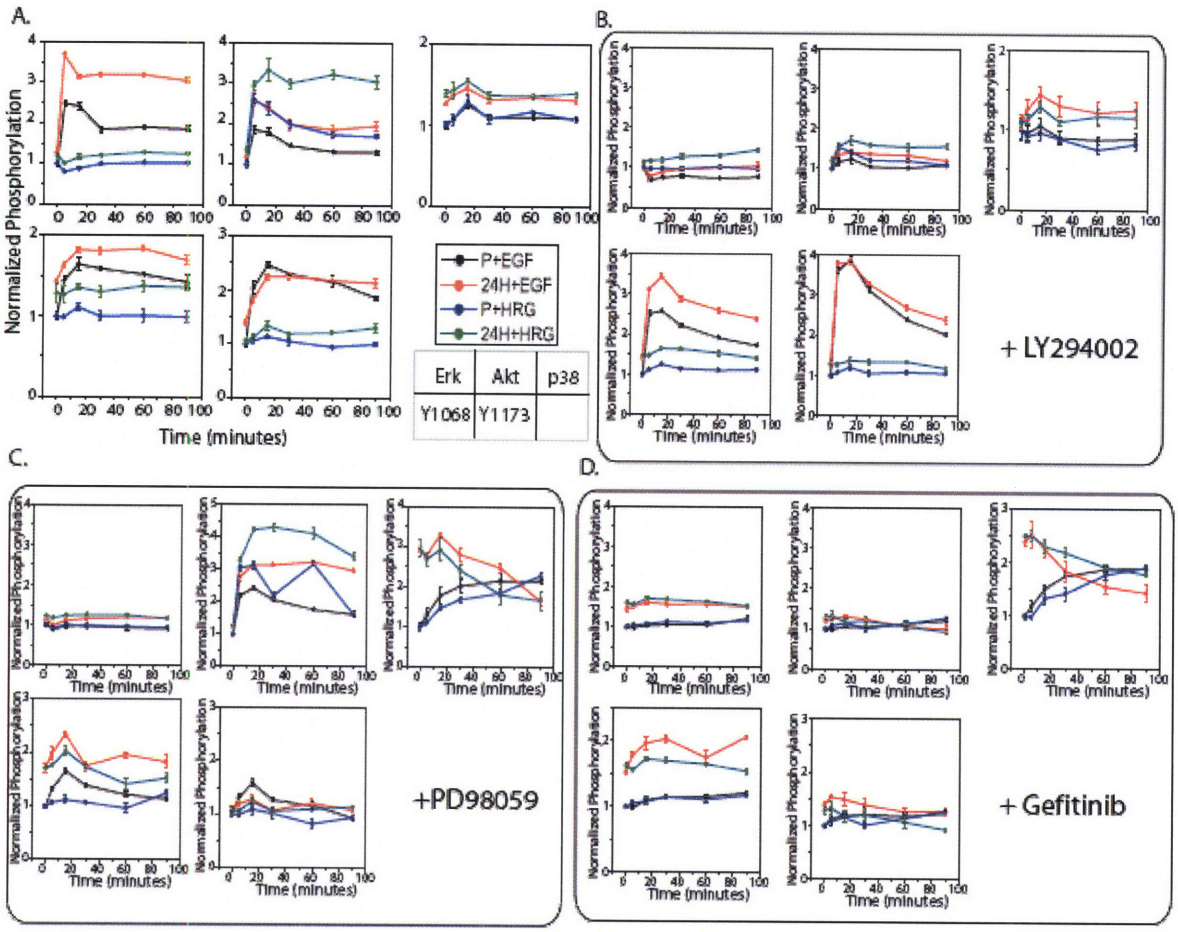


Figure 6-7: Effects of LY294002, PD98059, and Gefitinib on the phosphorylation of Erk, Akt, p38, and EGFR. Quantified Erk, Akt, p38, EGFR Y1068 and EGFR Y1173 phosphorylation for parental and 24H cells stimulated with EGF (100 ng/ml) or HRG (80 ng/ml) in the absence of inhibitor (A, as shown in Figure 6-3), pretreated with LY294002 (20 μM, B), pretreated with PD98059 (25μM, C), or pretreated with Gefitinib (20 μM, D). All data is shown ±SEM.

In sum, signal profiling in response to inhibition revealed novel connections between the signaling nodes measured and offered a new lens through which to understand the effects of individual inhibitors on cell migration.

6.3.5 A computational linear model creates intuition about how the network integrates the signals to regulate cell migration

To codify our intuition about how signals integrate to govern cell migration, we surmised that cell speed and persistence were functions of all five phosphorylation sites measured [i.e. persistence = f(pEGFR, pAkt, pp38, pErk)]. Although the exact nature of this function is probably very complicated, we approximated it as linear to arrive at the following general equation for persistence (we use persistence in the example here, but the same approach is applicable for speed):

$$persistence = \sum_i c_i x_i \quad (1)$$

where x_i refers to signaling metric i (e.g. normalized level of phosphorylated Akt at 5 minutes or p38 at 90 minutes) and c_i is a constant multiplier of the signal determining its weight in the final calculation of persistence. The vector c maps our signals (x) onto persistence, and we would obviously like to define a c such that we obtain an accurate prediction of persistence across a wide range of x values. Mathematically speaking, we define the vector c such that the range of the following linear transformation is large:

$$\mathbf{P} = \mathbf{X}c \quad (2)$$

where **P** is now a column vector (*Mx1*) of persistence values corresponding to *N* cellular conditions, **X** is the signaling matrix (*MxN*) comprised of *M* row vectors x, and c is a vector of constants (*Nx1*) that maps **X** onto **P**. The definition of c such that the range of

the linear transformation is large indicates that given diverse sets of signaling profiles, we can accurately predict cell persistence *a priori*. Of course, if there are conditions in which the vector *c* fails to map **X** accurately onto **P**, we know that we need to refine *c*, by either adding more signaling metrics to our original problem (i.e. profiling other signals that may be predictive of persistence) or adding more cellular conditions before we define the vector *c*.

So the exercise of solving for *c* is basically the codification of intuition about how each of our signals affects cell migration, as long as we limit our intuition to linear rules. For example, the model does not account for any intuition that we might gain from exploring how the product of pEGFR1173 and pAKT predict persistence. We can, however, include those terms explicitly in our model, but have not done so as exploration of these terms in the model revealed that they do not add to goodness of fit or prediction (data not shown).

Solving equation 2 using all cellular conditions available (e.g. parental and 24H data in the presence and absence of inhibitors), we defined the limit of how well linear reasoning can fit and predict cell persistence on the basis of our measured signals. We found that a linear model (the full model, see Materials and Methods for computational details) based on this data did strongly correlate with measured persistence values (R^2 = 0.89, R = 0.94, Figure 6-9B). The high goodness of fit and correlation values indicated the applicability of a linear approach to our dataset. Next, we sought to define a set of rules using only signaling and migration data in the absence of inhibitors and then explore whether these rules could be used to predict changes in persistence under various inhibitor treatments.

Using data obtained without inhibitors, we constructed a linear model for persistence (the reduced model). Using this equation and the experimentally measured inhibitor signaling sets, we predicted values of persistence in the presence of LY294002, PD98059, and Gefitinib. We found that *a priori* model predictions positively correlated with measured values of persistence (R = 0.80, Figure 6-9B). By contrast, prediction of persistence based on the measurement of any one kinase yielded an average correlation of 0.04 with a high of 0.35, indicating that signal state measurements are required for the accurate prediction of persistence in response to inhibition. In particular, the previously counter-intuitive effects of LY294002 on persistence (Figure 6-2), are *a priori* predicted by the reduced model (Figure 6-8).

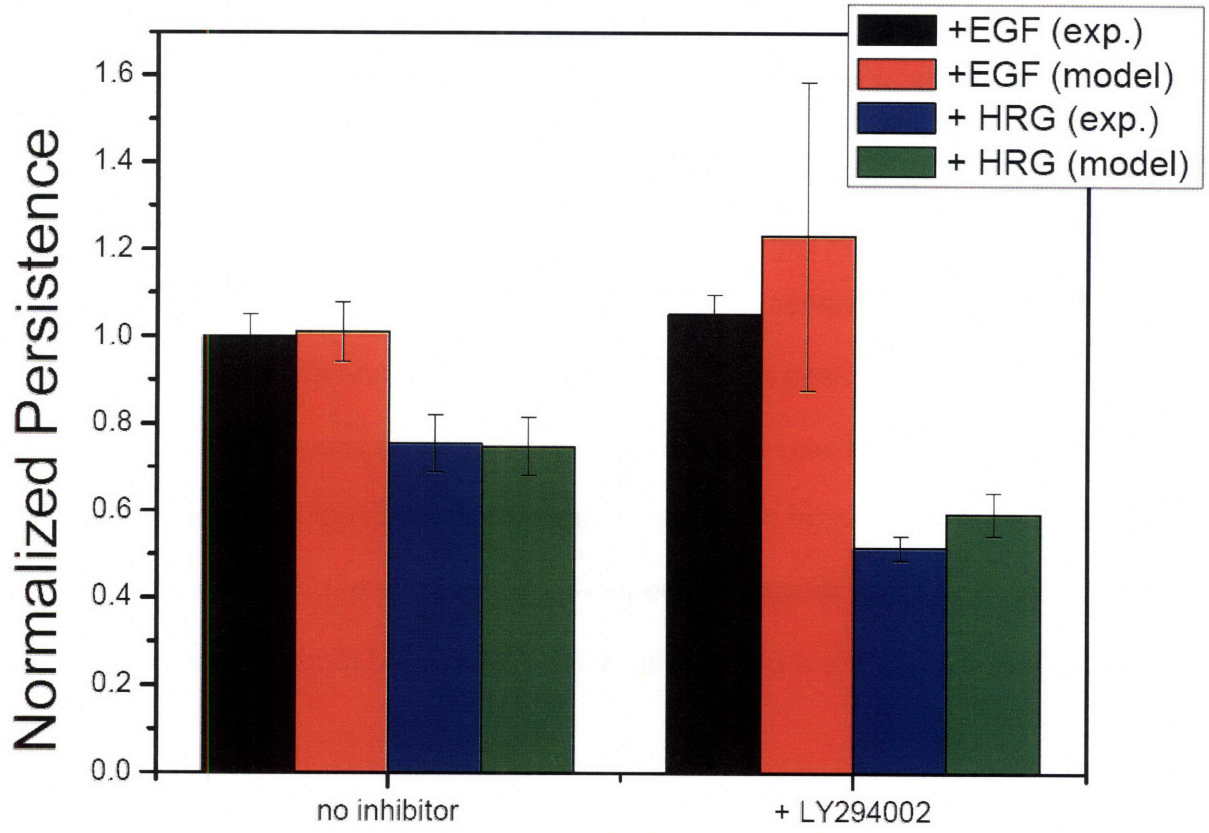


Figure 6-8: A reduced model predicts *a priori* the ligand-dependent effects of LY294002 inhibition on cell persistence in 24H cells. Experimental measurements of persistence of 24H cells treated with either EGF (100 ng/ml) or HRG (80 ng/ml) in the presence or absence of pretreatment with LY294002 (20 μM) are compared here to reduced model predictions of persistence values. Experimental data is shown ±SEM. Computational model data is shown ±SEM, where the error bars were obtained from jack-knifing.

Further inspection of the signaling sets obtained under inhibitor treatment revealed that phosphorylated levels of p38 did not vary significantly in the absence of inhibitors, but in the presence of PD98059 and Gefitinib, serum free levels of 24H p38 phosphorylation were significantly raised and phosphorylation dynamics were very different with 24H phosphorylation levels dropping and parental levels rising throughout the time course. We hypothesized that due to the lack of variance in p38

phosphorylation in the reduced model's input dataset, the model was overestimating levels of persistence for 24H serum-free cells and parental cells under PD98059 and Gefitinib treatment due to the higher 24H serum-free p38 phosphorylation levels and the rising parental levels. Thus, inclusion of additional p38 data to our reduced model was necessary to accurately capture the effects of p38 on persistence. Since the PD98059 signaling dataset had varied p38 regulation, we asked whether inclusion of this data would lead to a model that could predict persistence after treatment with Gefitinib and LY294002. We found that a new model based on singling data from cells exposed to no inhibitor and PD98059 accurately predicted persistence data (R = 0.93) almost as well as the full model. This increased accuracy was due to a revision of p38's role in persistence, with the reduced model assigning a positive role for p38 whereas the reduced model plus PD98058 data assigned it a negative role. Thus, the failings of our reduced model prompted the further inspection of p38 signaling which lead to a revised hypothesis of how p38 affects persistence.

The efficacy of inhibitors against the increased persistence measured in HER2 overexpressing cells was accurately captured by the full linear model under both EGF and HRG stimulation (Figure 6-9C,D). In the case of HRG stimulation, the model accurately captured the ineffectiveness of treatment with PD98059, and the increased efficacy of Gefitinib over LY294002 (Figure 6-9C). Notably, the model not only captured trends, but also accurately recapitulated absolute levels of persistence. In the case of EGF stimulation, the model captures the trend of increasing efficacy from LY294002 to PD98059 to Gefitinib (Figure 6-9D). In the case of Gefitinib, however, the model failed to capture the absolute amount of persistence. Nevertheless, the model, by accurately

recapitulating measured persistence levels, defined a set of rules based on signaling data that are predictive of persistence. The coefficients associated with each phosphorylation metric in the model reveal the metric's importance to persistence. Coefficient values, plotted in Figure 6-9E, indicated that Akt most positively correlates with persistence of the signals measured. However, EGFR Y1173 and early-phase Y1068 data also strongly correlate with persistence, indicating that an increase in phosphorylation at those sites is sufficient to keep persistence high even in the absence of Akt. Erk plays a positive but limited role in stimulating persistence, and p38 negatively correlates with persistence, indicating that decreasing p38 phosphorylation leads to increased directional persistence in the absence of any other change. Thus, the construction of a linear model for persistence defined a set of rules that integrate signal state information to describe changes in persistence in response to varying HER2 expression, ligand stimulation, and inhibitor treatment.

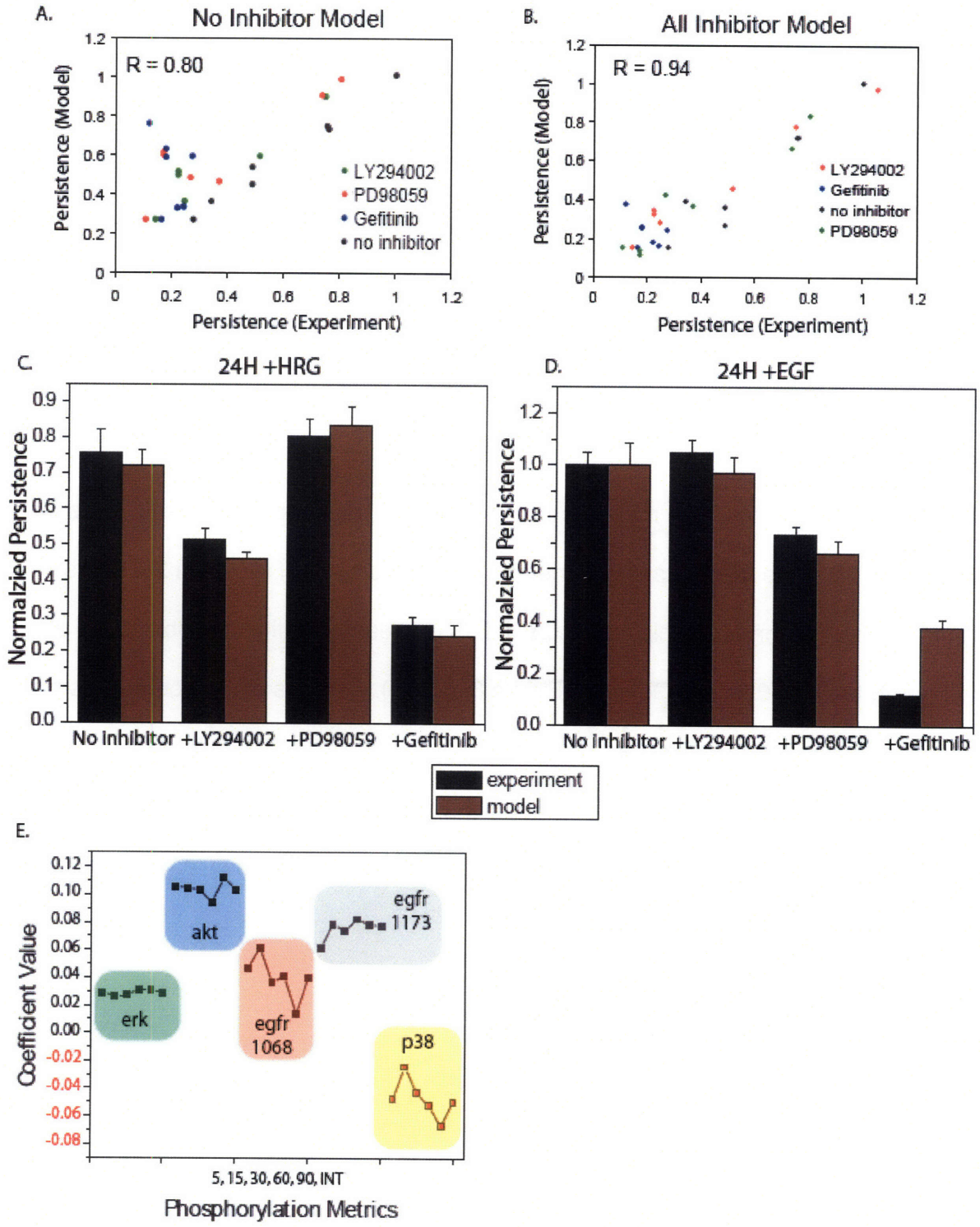


Figure 6-9: A linear model captures changes in persistence and identifies quantitative roles for phosphorylation in control of persistence. (A) Correlation between predictions from a linear model constructed only from signaling and persistence data in the absence of inhibitor and experimentally measured values of

persistence in the presence and absence of pretreatment with LY294002, PD98059, and Gefitinib. (B) Correlation between predictions from a linear model constructed from all inhibitor data and experimentally measured values of persistence. (C) Persistence predictions of the model generated from all available data as compared to experimental measurements for 24H cells stimulated with HRG (80 ng/ml) and (D) 24H cells stimulated with EGF (100 ng/ml). (E) Coefficient values for all phosphorylation metrics included in the full model. For each kinase, phosphorylation at 5, 15, 30, 60, 90, and the integral of phosphorylation are represented. Data in C and D are shown ±SEM.

A model constructed to predict changes in cell speed on the basis of only data without inhibitors predicted values for speed under inhibition that correlated positively with measured values (R = 0.77, Figure 6-10A). A model constructed from all available data correlated strongly with measured data (R = 0.87) and captured observed trends (Figure 6-10B). The model captured the significant decrease of speed in 24H cells treated with inhibitor and explained LY294002's decreased ability to inhibit speed as compared to PD98059 and Gefitinib (Figure 6-10C,D). Analysis of the coefficients from the full model reveal that Erk, early-phase Akt, and late-phase EGFR Y1173 phosphorylation correlate positively and strongly with speed, whereas EGFR Y1068 plays a diminished role in the governance of speed and p38 inversely correlates with increases in cell speed (Figure 6-10E).

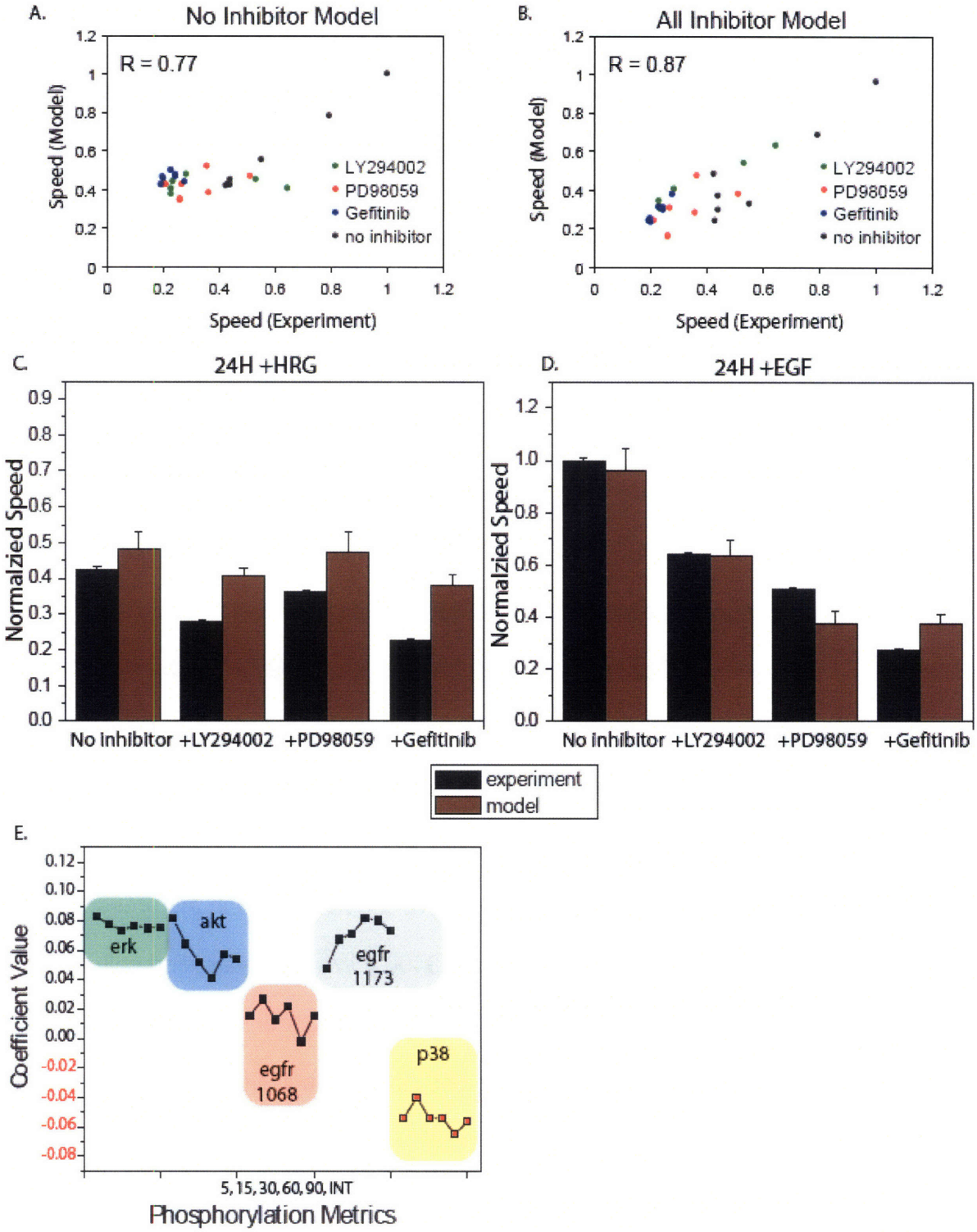


Figure 6-10: A linear model captures changes in speed and identifies quantitative role for phosphorylation in control of speed. (A) Correlation between predictions from a linear model constructed only from signaling and speed data in the absence of

inhibitor and experimentally measured values of speed in the presence and absence of pretreatment with LY294002, PD98059, and Gefitinib. (B) Correlation between predictions from a linear model constructed from all inhibitor data and experimentally measured values of speed. (C) Speed predictions of the model generated from all available data as compared to experimental measurements for 24H cells stimulated with HRG (80 ng/ml) and 24H cells stimulated with EGF (100 ng/ml, D). (E) Coefficient values for all phosphorylation metrics included in the full model. For each kinase, phosphorylation at 5, 15, 30, 60, 90, and the integral of phosphorylation are represented. Data in C and D are shown ±SEM.

6.4 Discussion and conclusions

In this study, we sought to understand the effects of treatment with small molecule inhibitors targeting EGFR, MEK, and PI3K on HER2-mediated cell migration. We surmised that although the inhibitors used had well-defined off-target binding properties [3, 18], the quantification of off-target effects due crosstalk in the signal state after inhibitor treatment would be critical to gain a consistent understanding of how drug treatment inhibits HER2-mediated directional persistence and speed.

We quantified the phosphorylation of four kinases and five phosphorylation sites in response to EGF and HRG under low or high HER2 conditions and in the presence or absence of Gefitinib (Iressa), PD98059, and LY294002. We also quantified cell migration, and in particular directional persistence, for cells under these same conditions. Analysis of the signaling dataset alone revealed significant off-target effects for each inhibitor, suggesting elements of crosstalk and feedback. Analysis of the migration dataset revealed ligand-dependent inhibition of persistence by some inhibitors and the generally high efficacy of Gefitinib as an inhibitor against HER2-mediated increases in directional persistence under both EGF and HRG stimulation conditions.

When viewed together, the migration and signaling datasets revealed the role of multiple upstream kinase phosphorylation events in the governance of persistent movement and the use of linear modeling codified these roles by ascribing a numerical importance to each signaling event. Using the model, the inhibition of persistence can be understood through an inhibitor's efficacy against its target and its effect on the greater signal state.

6.4.1 Signal measurements reveal crosstalk

Treatment of HMEC cells with LY294002 decreased Akt phosphorylation, notably in 24H cells stimulated with HRG, but also stimulated a greater amount of phosphorylation at EGFR Y1068 and Y1173 in EGF-treated cells as compared to cells not treated with inhibitor (Figure 6-7 A,B). Since trends in the differences (or lack thereof) of EGFR phosphorylation between parental and 24H cells were conserved when cells were treated with either no inhibitor or LY294002, these data indicate the possibility of negative feedback between PI3K and EGFR phosphorylation that is unaffected by HER2 expression levels in HMEC cells.

Treatment with LY294002 also interestingly prevented the phosphorylation of Erk, suggesting that PI3K acts upstream of Erk phosphorylation (Figure 6-7B). Similar results were obtained with Wortmannin, further suggesting that PI3K inhibition, and not off-target binding effects from LY294002, are responsible for the decreased Erk phosphorylation (Figure 6-5). Although the documented inhibition of Raf signaling by Akt [24] lead us to hypothesize that LY294002 would increase Erk phosphorylation, PI3K has been shown to regulate Erk phosphorylation in some cell lines, typically, however,

at low but not high EGF concentrations [25]. Nevertheless, our data show that even in the presence of saturating levels of EGF, treatment with LY294002 can result in the complete ablation of Erk phosphorylation. Furthermore, in the MCF10A breast epithelial cell line transfected with mutant H-Ras, inhibition of PI3K with LY294002 has been shown to inhibit Erk activation but not that of MEK, suggesting that PI3K regulates Erk activity through a MEK-independent pathway [12]. We observed, however, a complete absence of Erk signaling after treatment with LY294002 or PD98059, suggesting that both inhibitors target a single pathway responsible for Erk phosphorylation. Thus, our data indicate that PI3K interferes with the MEK-dependent phosphorylation of Erk. An interesting potential facilitator for PI3K's regulation of MEK may be Raf-1, which has been shown to dependent on PI3K activity in L6 cells, but has also been shown to be activated by PI3K inhibitors in other cell lines due to the afore mentioned negative regulation of Raf by Akt [24, 26].

Treatment with PD98059 and Gefitinib had very similar effects on p38 activity, with both inhibitors eventually decreasing the initially high serum-free phosphorylation levels in 24H cells to approximately the level observed without inhibitor after 90 minutes (Figure 6-7C,D). Concomitant increase in p38 phosphorylation was observed in parental cells, suggesting HER2-dependent differences in the signaling mechanisms governing the phosphorylation of p38. Treatment with PD98059 also dramatically decreased phosphorylation at EGFR Y1173 (Figure 6-7C). Previous studies have shown that phosphorylation at Y1173 recruits SHP-1, a phosphatase, that leads to dephosphorylation of EGFR and the attenuation of Erk phosphorylation [27]. Our data

suggest that there exists a feedback loop such that attenuation of Erk signaling serves to decrease phosphorylation at Y1173.

Treatment with Gefitinib removed phosphorylation of Akt in response to HRG, a result consistent with its documented ability to trap HER2 and HER3 receptors in an inactive state even though it is specific for the EGFR ATP domain [[28], Figure 6-7D]. Interestingly, we found that Gefitinib differentially knocks down EGFR Y1068 phosphorylation depending on HER2 expression levels, with EGF-stimulated 24H phosphorylation similar in the presence or absence of Gefitinib, while parental phosphorylation is significantly decreased in response to Gefitinib treatment (Figure 6-7D).

In sum, our signaling data indicate that given our current state of knowledge and a high degree of connectivity in signaling networks, it is very difficult to *a priori* predict the effect of inhibitors on a signal state, thus requiring signal profiling across key kinases to gain an accurate understating of inhibitor action at the signal and cell behavior levels.

6.4.2 Migration studies reveal ligand and HER2 dependent inhibition

Migration studies identified Gefitinib as a potent inhibitor of persistence and speed for both HRG and EGF stimulated 24H cells, suggesting the potential for Gefitinib's use as an anti-metastatic treatment in HER2 overexpressing tumors (Figure 6-4E,F,G). Interestingly, our data also revealed that PD98059 treatment was more effective in decreasing persistence in parental cells as compared to 24H cells, indicating

that HER2 overexpressing cells are less dependent on Erk for directionally persistent motion (Figure 6-4G). On the other hand, Gefitinib was more effective against persistence in 24H cells and nearly equalized values of persistence across cell type indicating that inhibition had resulted in the lowering of persistence to a baseline level independent of HER-signaling (Figure 6-4F,G). Finally, differences in the inhibitory efficacy of treatment with PD98059 and LY492004 depending on whether EGF or HRG was present revealed the importance of signal state context for understanding the role of one kinase target on cell migration (Figure 6-4G).

6.4.3 Linear modeling defines a logical framework within which kinase contribution to cell speed and persistence is understood

A reduced linear model, based only on signal state data obtained without inhibitor, proved that incorporation of network data is essential for explaining seemingly counter-intuitive effects of inhibition. In particular, although the finding that Akt inhibition did not have a consistent effect of persistence (Figure 6-2) initially surprised us, we were able to predict the ligand-dependent efficacy of inhibition *a priori* when we constructed a linear model for persistence based on signaling data from all five phosphorylation sites measured (Figure 6-8). Importantly, *a priori* predictions of persistence based on any one kinase fail to capture changes under inhibition, indicating the need for signal state measurement.

Our full linear model revealed self-consistent rules for the effect of each kinase studied on cell speed and persistence. Our results indicate that Akt phosphorylation

correlates positively with persistence (Figure 6-9). Akt's role in breast cancer migration is not clear, with recent reports indicating that Akt may act an inhibitor of metastasis in breast cancer tumors [29], while others reveal that Akt increases invasion and metastasis [10]. Our results indicate that Akt inhibition via LY294002 treatment can lead to an effective decrease of HER2-mediated cell migration in the absence of off-target effects such as increased EGFR phosphorylation. Thus, our data highlight the importance of cellular context for efficacy of inhibition. Since both HRG and EGF-family ligands have been found to circulate in breast cancer tumors [30], we speculate that inhibition of Akt with a PI3K inhibitor similar to LY294002 would be effective against cell migration in tumors with high levels of HRG but not necessarily in those with high EGF-family ligand expression, due to LY294002's role in increasing EGFR phosphorylation. In addition, our data contrast an earlier report suggesting that PI3K does not play a role in directed migration in the absence of chemical gradients [31]. Our data also suggest a positive role for EGFR phosphorylation in the governance of persistence, and in particular, Y1173 phosphorylation. Furthermore, EGFR is able to regulate persistence through a pathway that is not dependent on PI3K, as suggested by EGFR's role in maintaining high persistence in cells treated with LY294002 and EGF.

We have previously observed, under slightly different conditions, that Y1173 negatively correlated with cell migration (see previous chapter on PLSR models for mass spectrometry analysis). The difference between our results is due to the following factors: 1. the addition of time points past 30 minutes in the present study showed that 24H phosphorylation is generally slightly more sustained than parental phosphorylation, 2. the trend of higher parental phosphorylation in response to EGF is much more

prominent in the mass spectrometry dataset, where pretreatment with high amounts of insulin occurred, 3. the testing of the current dataset via inhibition added more cellular conditions for regression, thus changing the parameter vector and indicating that the coefficients derived for Y1173 from the mass spectrometry dataset might not be applicable for prediction across a large range of inhibitor data. The construction of a model on the basis of no inhibitor data and only time points equaling 30 minutes or less yields a negative coefficient for Y1173 at five minutes (although the value is very close to zero), indicating a consistency of measurement between the two studies for early time data.

Erk plays a positive but diminished role in directional persistence across the many conditions we sampled. This is due to the fact that Erk phosphorylation was ablated under all inhibitor treatments, while persistence values varied and sometimes remained high. This conclusion corroborates the fact that Erk, while linked to chemotaxis in a small number of cell systems, is generally considered less important than Akt for directional migration.

Our finding that p38 negatively correlates with directional persistence results from the effects of Gefitinib and PD98059 on p38 phosphorylation levels and the fact that neither ligand treatment stimulated a great deal of p38 phosphorylation. The reduced model predicted a positive role for p38 in the governance of cell persistence, but this is one of the reasons for its poor correlation to experimental data as compared to the full model. Our data contrasts literature indicating that p38 plays a positive role in the migration in breast epithelial MCF10A cells [12], indicating that p38 may play a different role in HMEC cells under HER2 overexpression. In addition, a recent study

shows that Rac mediates increases in p38 phosphorylation that in turn linked increase in cell migration [12]. On the other hand, decreased Rac phosphorylation has been shown to increase directional persistence in the absence of chemical gradients, suggesting that p38 phosphorylation may be a proxy for decreased Rac activity responsible for increased cell persistence [31].

Inspection of the full models for speed and persistence indicate that similar upstream kinases both positively regulate speed and persistence, but with different importance. For instance, Erk plays a more prominent role in the governance of cell speed as compared to persistence. In contrast late-phase Akt signaling plays a more prominent role in the governance of cell persistence as compared to speed, as evidenced by the high sustained Akt levels in 24H cells stimulated with HRG.

6.4.4 Conclusions

We have shown that the efficacy of any inhibitor on HER2-mediated changes in directional persistence is a function of its effect on the network signal state and not simply its target. By compiling signal state and cell migration data in response to three separate inhibitors we showed how initially surprising results based on an understanding of directional persistence in the context of one kinase (Figure 6-2), can be explained through the incorporation of signal state measurements and linear modeling (Figure 6-9). In the case presented in Figure 6-2, we found that LY294002 does not inhibit migration in EGF-stimulated 24H cells due to an off-target increase in EGFR phosphorylation. With the already heavy focus on the identification of off-target

binding effects through biochemical assays [3], and the advent of new high-throughput technologies that allow for rapid and more exhaustive characterization of signal state in response to many inhibitors [2], we anticipate that the approach highlighted here will be further applied in new biological systems to achieve a better understanding of how to control cell behavior by drugging with small molecule inhibitors.

References

1. 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.
2. Sevecka, M. and G. Macbeath, *State-based discovery: a multidimensional screen for small-molecule modulators of EGF signaling.* Nat Methods, 2006. **3**(10): p. 825-31.
3. Fabian, M.A., et al., *A small molecule-kinase interaction map for clinical kinase inhibitors.* Nat Biotechnol, 2005. **23**(3): p. 329-36.
4. Hynes, N.E. and D.F. Stern, *The biology of erbB-2/neu/HER-2 and its role in cancer.* Biochim Biophys Acta, 1994. **1198**(2-3): p. 165-84.
5. Harari, D. and Y. Yarden, *Molecular mechanisms underlying ErbB2/HER2 action in breast cancer.* Oncogene, 2000. **19**(53): p. 6102-14.
6. Kumar, N., et al., *A high-throughput migration assay reveals HER2-mediated cell migration arising from increased directional persistence.* Biophys J, 2006. **91**(4): p. L32-4.
7. Spencer, K.S., et al., *ErbB2 is necessary for induction of carcinoma cell invasion by ErbB family receptor tyrosine kinases.* J Cell Biol, 2000. **148**(2): p. 385-97.

8. Deisboeck, T.S., T. Demuth, and Y. Mansury, *Correlating velocity patterns with spatial dynamics in glioma cell migration.* Acta Biotheor, 2005. **53**(3): p. 181-90.

9. Friedl, P., et al., *Migration of coordinated cell clusters in mesenchymal and epithelial cancer explants in vitro.* Cancer Res, 1995. **55**(20): p. 4557-60.

10. 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.

11. Hirsch, D.S., Y. Shen, and W.J. Wu, *Growth and motility inhibition of breast cancer cells by epidermal growth factor receptor degradation is correlated with inactivation of Cdc42.* Cancer Res, 2006. **66**(7): p. 3523-30.

12. Shin, I., et al., *H-Ras-specific activation of Rac-MKK3/6-p38 pathway: its critical role in invasion and migration of breast epithelial cells.* J Biol Chem, 2005. **280**(15): p. 14675-83.

13. Ho, W., et al., *A differential role of extracellular signal-regulated kinase in stimulated PC12 pheochromocytoma cell movement.* Exp Cell Res, 2001. **263**(2): p. 254-64.

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

15. Maheshwari, G., H.S. Wiley, and D.A. Lauffenburger, *Autocrine epidermal growth factor signaling stimulates directionally persistent mammary epithelial cell migration.* J Cell Biol, 2001. **155**(7): p. 1123-8.

16. Merlot, S. and R.A. Firtel, *Leading the way: Directional sensing through phosphatidylinositol 3-kinase and other signaling pathways.* J Cell Sci, 2003. **116**(Pt 17): p. 3471-8.

17. Oda, K., et al., *A comprehensive pathway map of epidermal growth factor receptor signaling.* Mol Syst Biol, 2005. **1**: p. 2005 0010.

18. Davies, S.P., et al., *Specificity and mechanism of action of some commonly used protein kinase inhibitors.* Biochem J, 2000. **351**(Pt 1): p. 95-105.

19. Wolf-Yadlin, A., et al., *Effects of HER2 overexpression on cell signaling networks governing proliferation and migration.* Mol Syst Biol, 2006. **2**: p. 54.

20. Band, V. and R. Sager, *Distinctive traits of normal and tumor-derived human mammary epithelial cells expressed in a medium that supports long-term growth of both cell types.* Proc Natl Acad Sci U S A, 1989. **86**(4): p. 1249-53.

21. Miller-Jensen, K., et al., *Adenoviral vector saturates Akt pro-survival signaling and blocks insulin-mediated rescue of tumornecrosis-factor-induced apoptosis.* J Cell Sci, 2006. **119**(Pt 18): p. 3788-98.

22. Eriksson, L., et al., *Multi- and Megavariate Data Analysis: Principles and Applications.* 2001, Umea, Sweden: Umetrics.

23. Schulze, W.X., L. Deng, and M. Mann, *Phosphotyrosine interactome of the ErbB-receptor kinase family.* Mol Syst Biol, 2005. **1**: p. 2005 0008.

24. Zimmermann, S. and K. Moelling, *Phosphorylation and regulation of Raf by Akt (protein kinase B).* Science, 1999. **286**(5445): p. 1741-4.

25. Wennstrom, S. and J. Downward, *Role of phosphoinositide 3-kinase in activation of ras and mitogen-activated protein kinase by epidermal growth factor.* Mol Cell Biol, 1999. **19**(6): p. 4279-88.

26. Cross, D.A., et al., *The inhibition of glycogen synthase kinase-3 by insulin or insulin-like growth factor 1 in the rat skeletal muscle cell line L6 is blocked by wortmannin, but not by rapamycin: evidence that wortmannin blocks activation of the mitogen-activated protein kinase pathway in L6 cells between Ras and Raf.* Biochem J, 1994. **303 (Pt 1)**: p. 21-6.

27. Keilhack, H., et al., *Phosphotyrosine 1173 mediates binding of the protein-tyrosine phosphatase SHP-1 to the epidermal growth factor receptor and attenuation of receptor signaling.* J Biol Chem, 1998. **273**(38): p. 24839-46.

28. Anido, J., et al., *ZD1839, a specific epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor, induces the formation of inactive EGFR/HER2 and EGFR/HER3 heterodimers and prevents heregulin signaling in HER2-overexpressing breast cancer cells.* Clin Cancer Res, 2003. **9**(4): p. 1274-83.

29. Yoeli-Lerner, M., et al., *Akt blocks breast cancer cell motility and invasion through the transcription factor NFAT.* Mol Cell, 2005. **20**(4): p. 539-50.

30. Normanno, N., et al., *Epidermal growth factor receptor (EGFR) signaling in cancer.* Gene, 2006. **366**(1): p. 2-16.

31. Pankov, R., et al., *A Rac switch regulates random versus directionally persistent cell migration.* J Cell Biol, 2005. **170**(5): p. 793-802.

Chapter 7 Quantitative analysis of Akt phosphorylation and activity in response to EGF and insulin treatment

Previous signaling work (Chapter 3) revealed potential differences between Akt phosphorylation levels and activity levels. In this chapter, we apply the measurement techniques introduced in Chapter 3 to understand how Akt activity is regulated through phosphorylation.

7.1 Introduction

The protein kinase Akt/PKB is a critical regulator of cellular functions such as apoptosis, proliferation, and migration [1, 2]. Its well-established role in the governance of cell survival implicates it as a critical signaling node in cancer, and its overexpression and increased activation has been found in a variety of cancers such as those occurring in the breast, neck, and lungs [3]. Prior studies into the mechanism of Akt's kinase activity have revealed that growth factor or insulin induced activation of phosphoinositide 3-kinase (PI3K) leads to the generation of 3,4,5 phosphatidylinositol (PIP3) and subsequent recruitment of Akt to the plasma membrane via its pleckstin homology (PH) domain. Once at the membrane, Akt is phosphorylated by phospoinositide-dependent kinase 1 (PDK1) on the threonine 308 residue (T308) that resides in its activation loop. In addition, phosphorylation on the serine 473 residue (S473) residing in its carboxy-terminal hydrophobic domain occurs via a kinase whose identity has been debated [4-6], although recent evidence identifies a Rictor-mTOR complex as the responsible kinase [7]. The T308 and S473 phosphorylation sites have been the primary focus of a large number of biochemical studies into Akt's mechanism

of activation. Salient results include the finding that phosphorylation on both the S473 and T308 sites is necessary for full kinase activation in response to insulin [with S473 phosphorylation alone inducing no activity and T308 phosphorylation alone inducing approximately one-third maximum activity [8]], that S473 phosphorylation precedes and promotes T308 phosphorylation [9], and that the dephosphorylation of the two sites often occurs differentially and through separate phosphatases [10, 11].

Although the majority of work regarding Akt activation has focused on phosphorylation at T308 and S473, recent evidence suggests the existence of other phosphorylation sites that regulate kinase activity. In particular, tyrosine phosphorylation has been identified as a key regulatory mechanism in Src-transformed cells, cells stimulated with epidermal growth factor (EGF), and cells treated with the tyrosine phosphatase inhibitor pervanadate [12, 13]. Threonine 72 and serine 246 have also recently been identified as autophosphorylated sites that regulate kinase activity [14]. Furthermore, uncoupling of T308 phosphorylation and kinase activity after initial kinase activation has been reported in response to insulin [11].

Given evidence that phosphorylation at multiple sites other than T308 and S473 are important for kinase activity, especially in systems highly relevant to cancer such as Src-transformed cell lines or those exposed to elevated levels of EGF ligands, we wondered how accurately inferences about kinase activity and its role in signaling could be made from interrogation of phosphorylation at S473 or T308 alone. The reason that these two sites, and in particular S473, have been profiled extensively in cell line and higher level tissue systems is the existence of phospho-specific antibodies against T308 and S473, making western blotting and immunostaining possible.

To answer this question, we quantified kinase activity in addition to T308 and S473 phosphorylation in two different cell lines, a Chinese hamster ovary cell line transfected with EGFR (CHO-EGFR) and a colon carcinoma cell line (HT-29), treated individually with EGF or insulin. We measured kinase activity over two hours to capture the initial kinase activation profile (with a time scale typically under 5 minutes) as well as deactivation or sustained activity at longer times, since both acute and longer term responses may be important for Akt's governance of cellular phenotype. Our results show that phosphorylation at T308 and S473 provides an accurate representation of kinase activity in the case of EGF stimulation in both cell lines. T308 and S473 phosphorylation, however, fail to capture elements of kinase activity in response to insulin treatment, most notably in the case of HT-29 cells where early phase oscillations in activity are not reflected by T308 or S473 phosphorylation levels. Additionally, our data suggest that both phosphorylation and dephosphorylation at T308 and S473 are tightly coupled under most conditions studied, in contrast to previous studies, and thus in these cases each site serves equally well as a proxy for kinase activity.

7.2 Materials and Methods

7.2.1 Cell culture and treatment

HT-29 cells (ATCC) were grown in McCoy's 5A medium supplemented with 10% fetal bovine serum, 2 mM glutamine, 100 units/ml penicillin, and 100 µg/ml streptomycin (Invitrogen). CHO K1 cells, transfected with EGFR-GFP as described previously [15], were grown in high-glucose Dulbecco's modified Eagle's medium containing 10% fetal

bovine serum, 2 mM glutamine, 1 mM sodium pyruvate, 1mM non-essential amino acids, 100 units/ml penicillin and 100 µg/ml streptomycin. The growth medium was supplemented with 500 µg/ml of G418 for plasmid expression maintenance and selectivity.

For lysis, cells were seeded at 50,000 cells/cm^2, grown for 48 hrs, and then stimulated with 100 ng/ml EGF (Peprotech) or 500 ng/ml insulin (Sigma) for the indicated times. Cells were lysed in 1% Triton X-100, 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 50 mM β-glycerophosphate, 20 mM sodium pyrophosphate, 30 mM NaF, 1 mM benzamidine, 2 mM EGTA, 200µM $NaVO_4$, 1 mM dithiothreitol (DTT), 1 mM phenylmethylsulfonyl fluoride, 10 µg/ml aprotinin, 10 µg/ml leupeptin, 10 µg/ml pepstatin, and 1 µg/ml microcystin-LR. Protein concentrations were determined with a micro bicinchoninic acid assay (Pierce).

7.2.2 Western blotting

To quantify phosphorylation levels, 80 µg of lysate were resuspended in 40 µl of sample buffer [100 mM DTT, 2% SDS, 10% glycerol, 0.01% bromophenol blue, 62.5 mM Tris-HCl (pH 6.8)]. Gel electrophoreses (10% polyacrylamide gel) was followed by transfer to polyvinylidene difluoride membranes (Biorad). Membranes were blocked with 5% nonfat milk or 5% bovine serum albumin in 20 mM Tris-HCl (pH 7.5), 137 mM NaCl, and 0.1% Tween-20. Membranes were then probed with anti-phospho-Akt (Ser473, #9271, Cell Signaling) or anti-phospho-Akt (Thr308, #4056, Cell Signaling). The membranes were then probed with horseradish-peroxidase-conjugated anti-rabbit secondary antibody (Amersham Pharmacia Biotech) and visualized by enhanced

chemiluminescence (Amersham Pharmacia Biotech) on a Kodak Image Station (Perkin Elmer). Densitometry was performed using molecular imaging software (Kodak). Band area net intensities were normalized to the 5 minute (for HT-29 cells) or 10 minute (for CHO-EGFR cells) value to produce the time series presented in Figures 7-5-7-8. Linearity for each antibody was established using serial dilutions of an insulin-stimulated 5 min. lysate from HT-29 cells (see Figures 7-1 and 7-2).

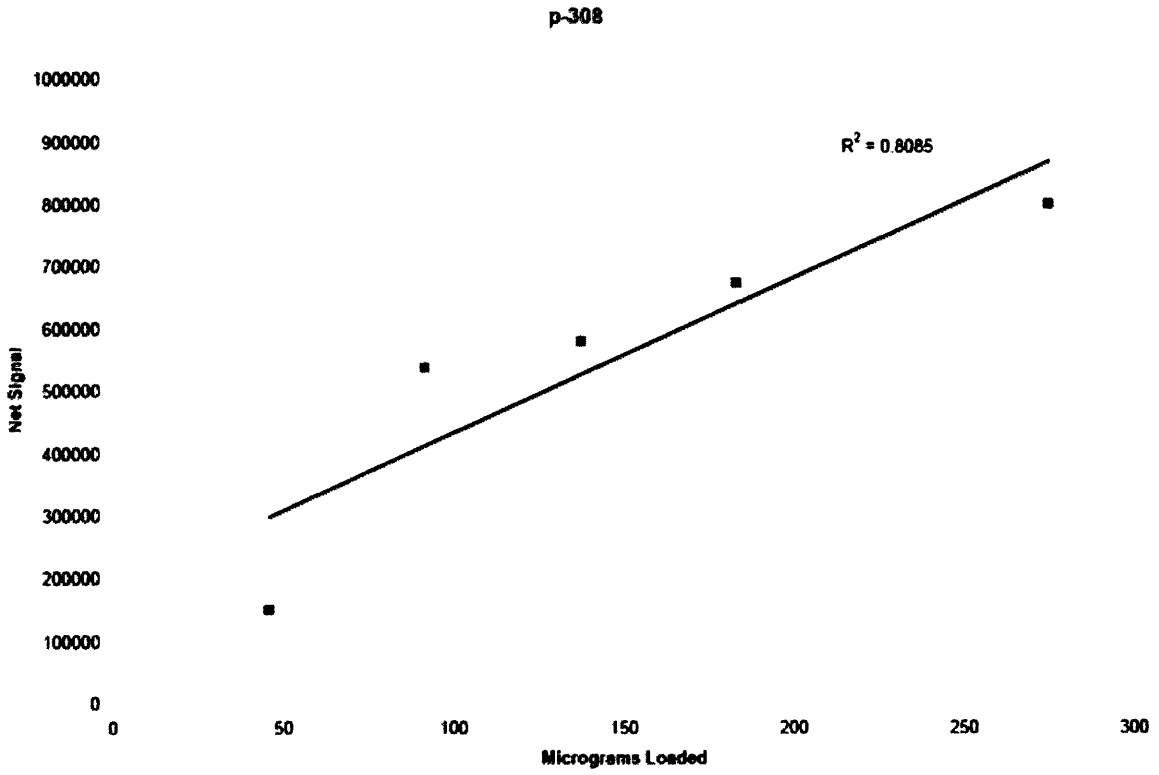


Figure 7-1: Linearity for T308 western blot. Using serially diluted lysates, we probed the linearity of densitometry as described in Materials and Methods.

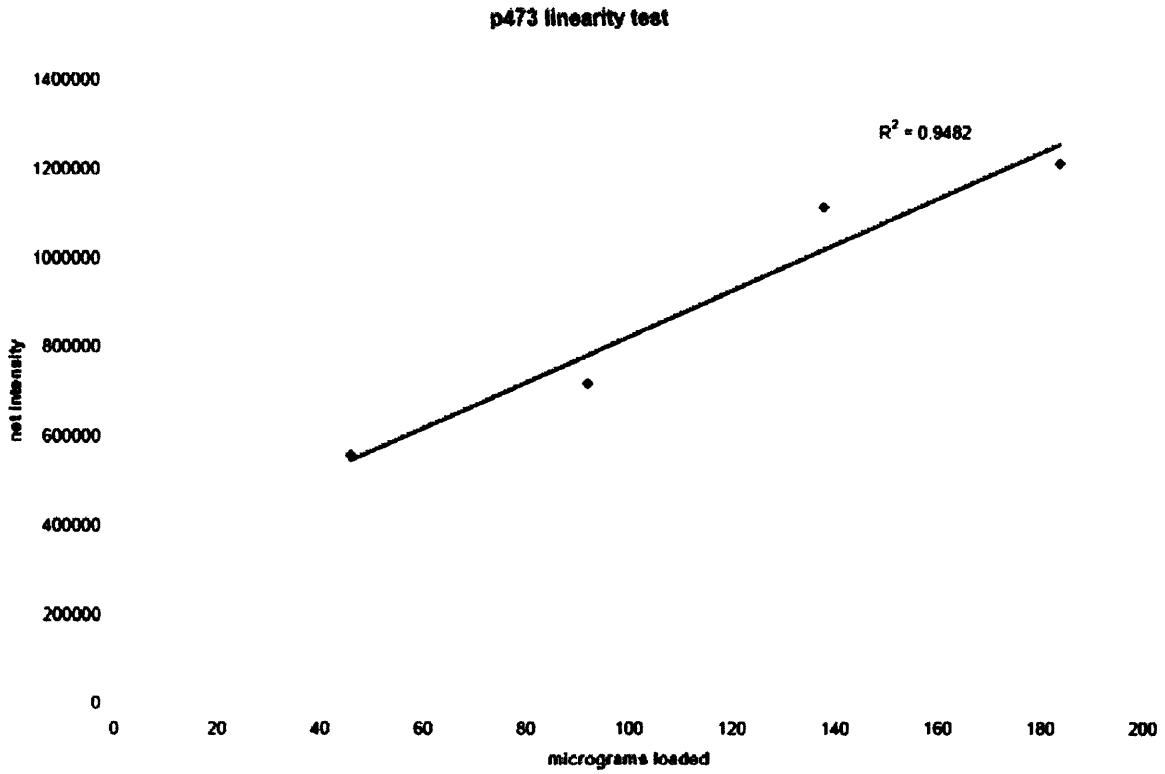


Figure 7-2: Linearity for S473 western blot. Using serially diluted lysates, we probed the linearity of densitometry as described in Materials and Methods.

7.2.3 Kinase activity assay

Kinase activity assays were performed as previously described [16]. Briefly, anti-Akt antibody (Upstate Biotech) was incubated in 96-well protein G-coated plates (Pierce) overnight. Lysates were then added and incubated overnight as well. Subsequent exposure to [γ^{32}-P]ATP and Aktide substrate initiated an *in vitro* reaction that was subsequently terminated after 30 minutes by addition of phosphoric acid. Reaction mixtures were then transferred to a phosphocellulose filter plate and filter bound [γ^{32}-P]-substrate was quantified using a scintillation counter. Linearity of the

assay in each cell type has been established in [16] and Figure 7-3. Count per minute readings were normalized to lysate concentrations and then to the 5 minute (for HT-29 cells) or 10 minute (for CHO-EGFR cells) value to produce the time series presented in Figures 7-5-8.

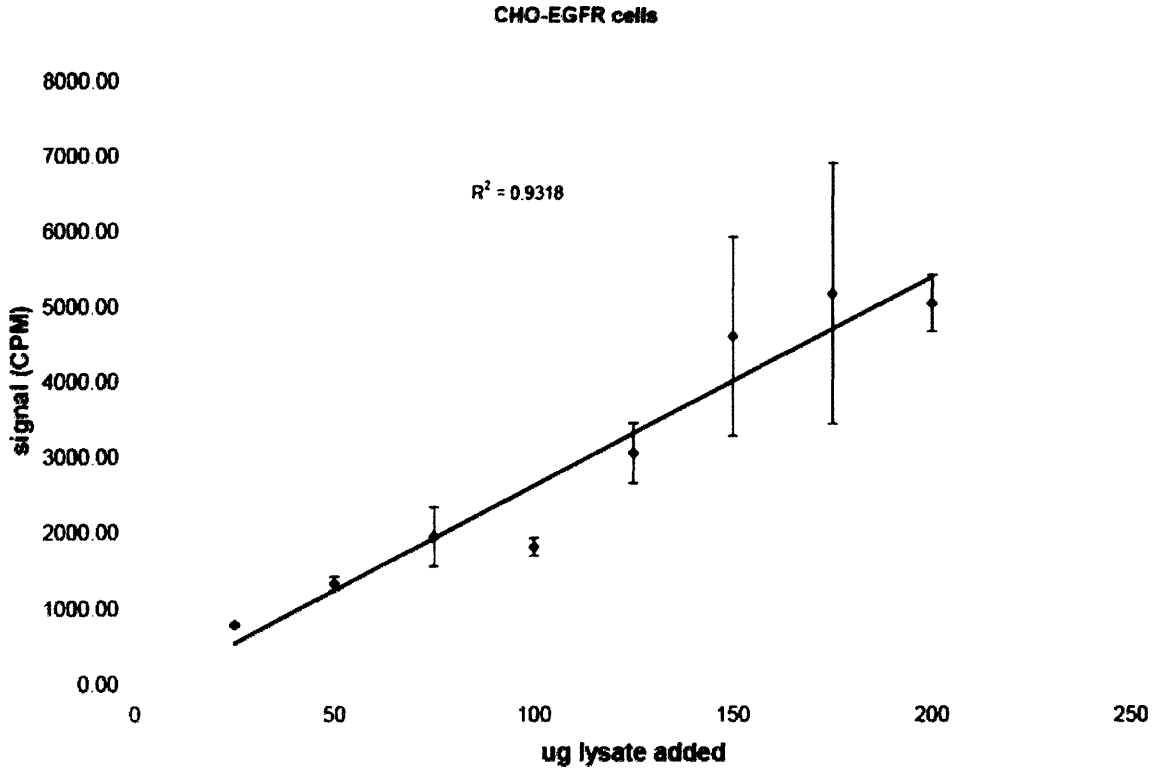


Figure 7-3: Linearity for Akt kinase activity assay in CHO-EGFR cells. Using serially diluted lysates, we probed the linearity of the Akt kinase activity assay as described in Materials and Methods.

7.2.4 Statistical analysis

Pearson correlation (R) values and p-values using student's t-test (95% confidence intervals) were obtained in Microsoft Excel.

7.3 Results

7.3.1 An experimental strategy for the quantitative comparison of Akt phosphorylation and activity

To directly compare phosphorylation and kinase activity, we conducted quantitative western blots (T308 and S473) and a kinase activity assay from individual lysates corresponding to one of three biological replicates for a particular cellular treatment (Figure 7-4). Each measurement technique was validated for linearity as described in the Methods section (Figures 7-1 and 7-2).

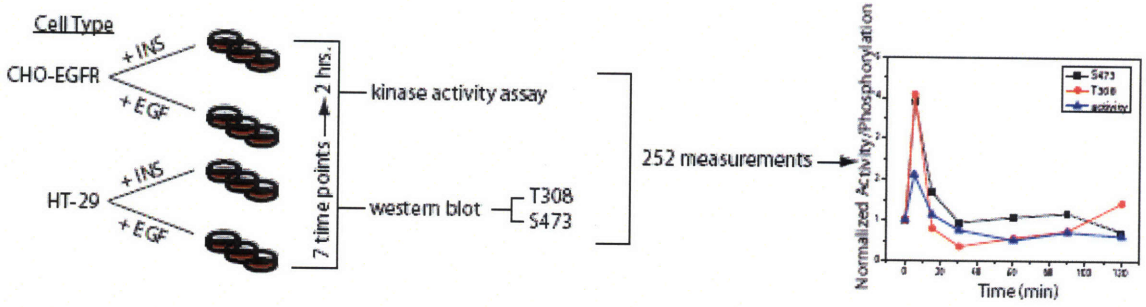


Figure 7-4: An experimental strategy to quantify and correlate Akt phosphorylation and kinase activity. Two cell types, CHO-EGFR and HT-29, were individually stimulated with EGF (100 ng/ml) or insulin (INS, 500 ng/ml). At each of seven time points distributed throughout two hours, three lysates (biological replicates) per treatment condition were generated. From each lysate, a kinase activity assay and two quantitative western blot against T308 and S473 were performed. Quantification

and normalization then allowed comparison between activity and phosphorylation time courses.

7.3.2 EGF treatment stimulates a transient Akt response in HT-29 cells and a sustained Akt response in CHO-EGFR cells

When HT-29 cells were treated with EGF (100 ng/ml), a transient ~3-fold activation was observed (Figure 7-5A). Quantification of T308 and S473 phosphorylation revealed a similar trend, with phosphorylation and subsequent dephosphorylation occurring rapidly within 15 minutes of ligand treatment (Figures 7-5B, C). The correlation between kinase activity and phosphorylation over the 2 hour time course was high, with $R \geq 0.95$ in both cases (Figure 7-5D). The phosphorylation and dephosphorylation trends for T308 and S473 correlated strongly with each other, yielding an $R = 0.96$ (Figure 7-5D). In contrast to HT-29 cells, CHO-EGFR cells treated with EGF exhibited sustained kinase activity that peaked after approximately 30 minutes (Figure 7-6A). Concomitant phosphorylation at the T308 and S473 was also observed (Figure 7-6B), as captured by the strong correlation between each site and kinase activity (Figure 7-6C, D). As was the case in the HT-29 cells, correlation between the two phosphorylation sites was high ($R = 0.94$, Figure 7-6D). Thus, the phosphorylation levels of T308 and S473 each accurately reflect kinase activity in two cell lines exhibiting unique temporal responses to EGF stimulation.

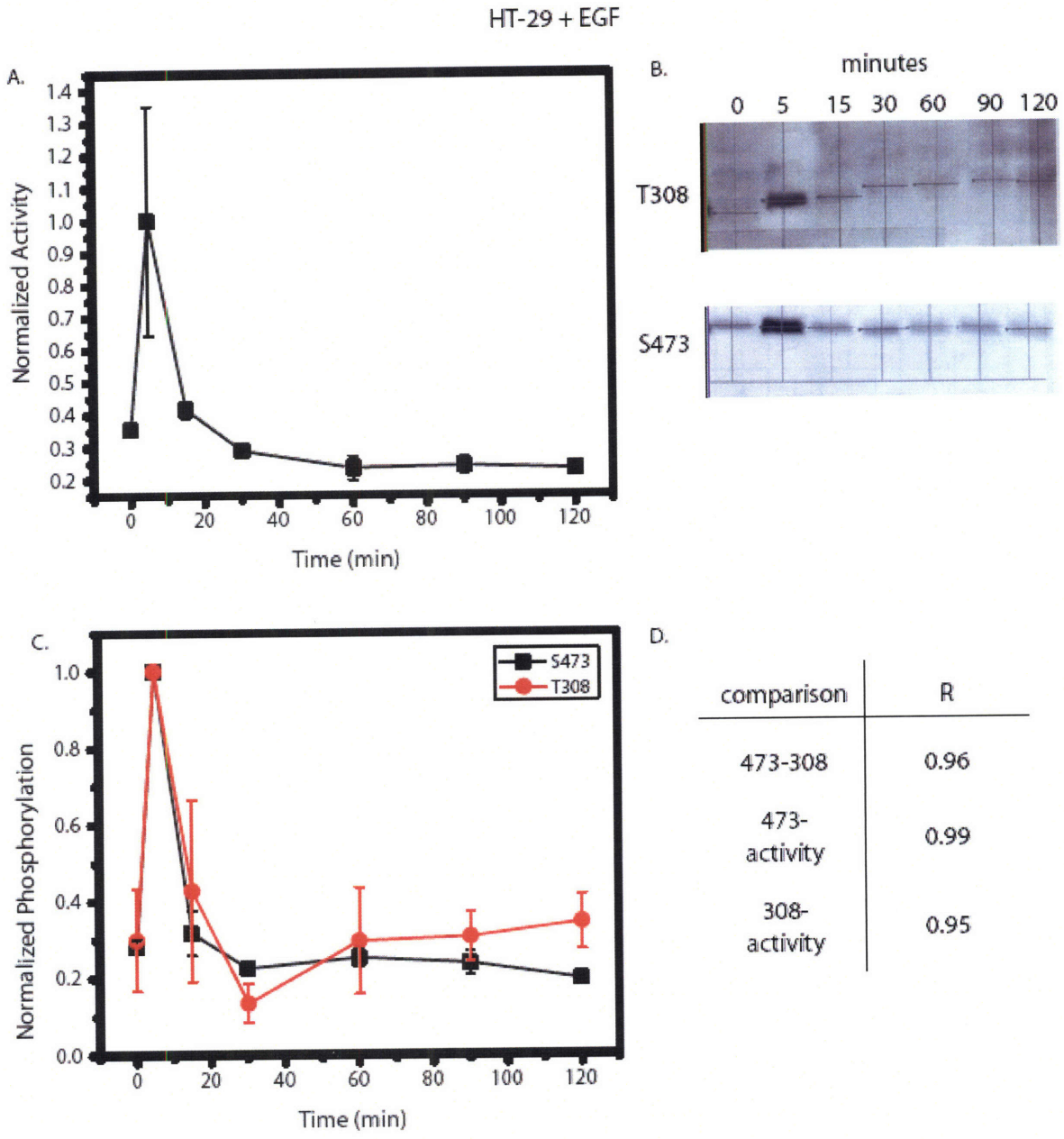


Figure 7-5: HT-29 cells treated with EGF exhibit transient Akt activation and phosphorylation. An in vitro kinase activity assay was used to measure Akt activity in HT-29 cells treated with EGF (100 ng/ml) at 0, 5, 15, 30, 60, 90, and 120 minutes (A). Phosphorylation at T308 and S473 was also measured under these conditions using western blot analysis. Shown in (B) are representative blots for T308 and S473 from the three biological replicates measured. Densitometry was used to quantify the net band intensity for all western blots (C). Calculation of the Pearson's correlation between phosphorylation of T308, S473, and kinase activity is shown in (D). All points in the time courses are the average of three biological replicates ± SEM. Time points were normalized to 5 minute kinase activity or phosphorylation levels.

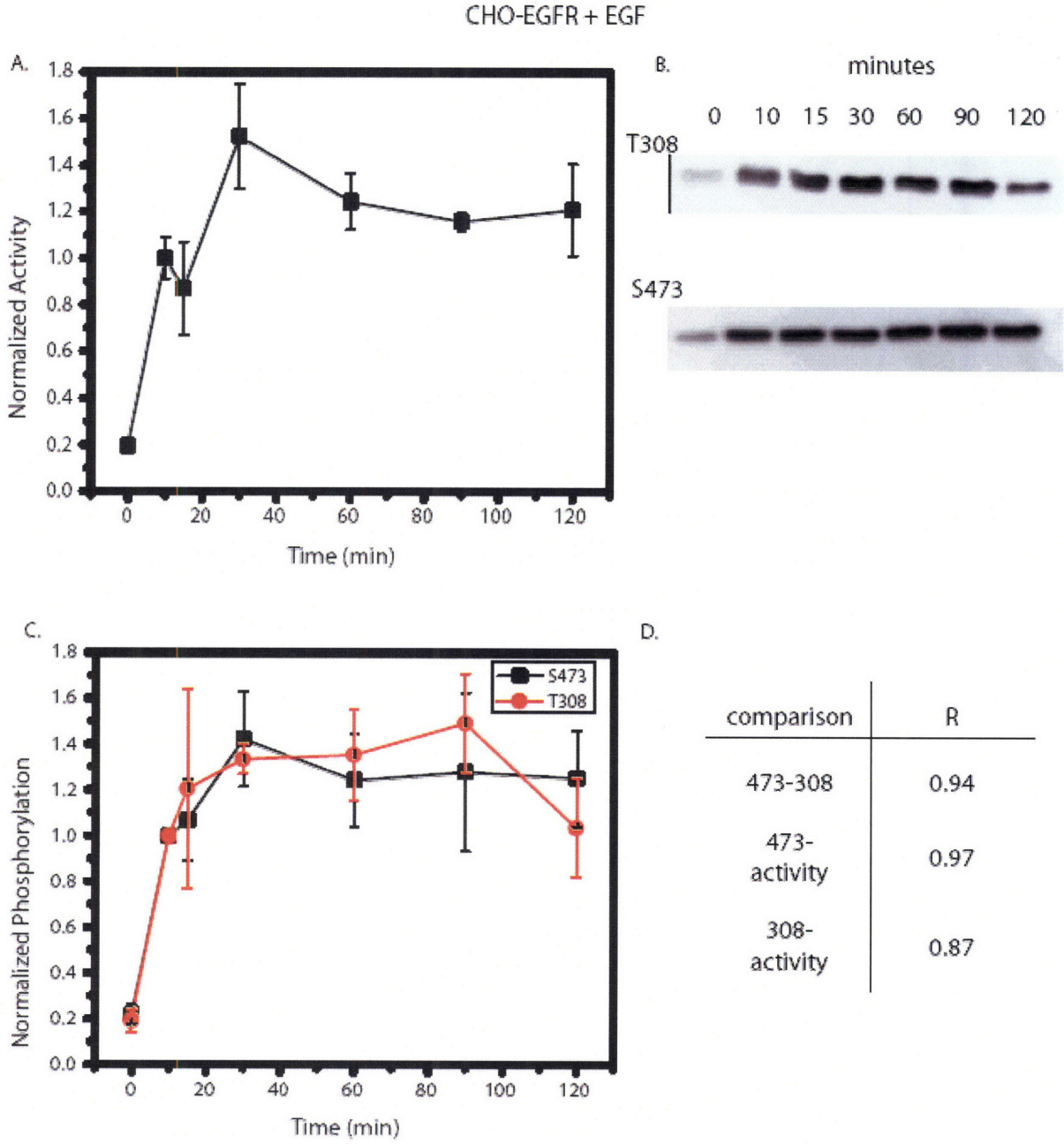


Figure 7-6: CHO-EGFR cells treated with EGF exhibit sustained Akt activation and phosphorylation. An in vitro kinase activity assay was used to measure Akt activity in CHO-EGFR cells treated with EGF (100 ng/ml) at 0, 10, 15, 30, 60, 90, and 120 minutes (A). Phosphorylation at T308 and S473 was also measured under these conditions using western blot analysis. Shown in (B) are representative blots for T308 and S473 from the three biological replicates measured. Densitometry was used to quantify the net band intensity for all western blots (C). Calculation of the Pearson's correlation between phosphorylation of T308, S473, and kinase activity is shown in (D). All points in

the time courses are the average of three biological replicates ± SEM. Time points were normalized to 10 minute kinase activity or phosphorylation levels.

7.3.3 Insulin treatment induces sustained AKT kinase activity in both HT-29 and CHO-EGFR cell lines that is not fully captured by T308 and S473 phosphorylation

HT-29 cells treated with insulin exhibited sustained Akt activity throughout the two hour time course. Interestingly, a statistically significant oscillatory behavior was observed, with the differences between subsequent time points from 5 to 60 minutes significant at $p < 0.05$ (Figure 7-7A). These oscillations were not reflected in the phosphorylation patterns of either S473 or T308 (Figure 7-7B, C), as captured by the low correlation between phosphorylation and activity [$R = 0.62$ and 0.57 for S473 and T308, respectively (Figure 7-7D)]. Despite low correlation with activity, the phosphorylation levels at T308 and S473 correlated strongly with each other, as reflected by the high correlation coefficient ($R = 0.91$, Figure 7-7D).

CHO-EGFR cells stimulated with insulin exhibited sustained Akt activity, but do not show any of the oscillatory behavior identified in the HT-29 cells (Figure 7-8A). Phosphorylation levels of both T308 and S473 reflect sustained kinase activity, although T308 phospho-levels in particular decline on average, a trend not seen in kinase activity. The correlation between the average phosphorylation measurements for T308 and S473 as shown in Figure 7-8C is relatively high ($R = 0.79$), but still significantly lower than seen in any of the other cellular conditions (all R's > 0.90), due to the previously mentioned T308 dephosphorylation trend that is not present for S473

phosphorylation. The S473 phosphorylation levels correlate strongly with activity (R = 0.93), whereas phosphorylation levels of T308 correlate weakly with kinase activity (R = 0.59). Thus, the measurement of S473 or T308 phosphorylation in response to insulin may not be enough to infer quantitative changes in kinase activity for HT-29 cells and only S473 phosphorylation correlates strongly with kinase activity in the CHO-EGFR cells.

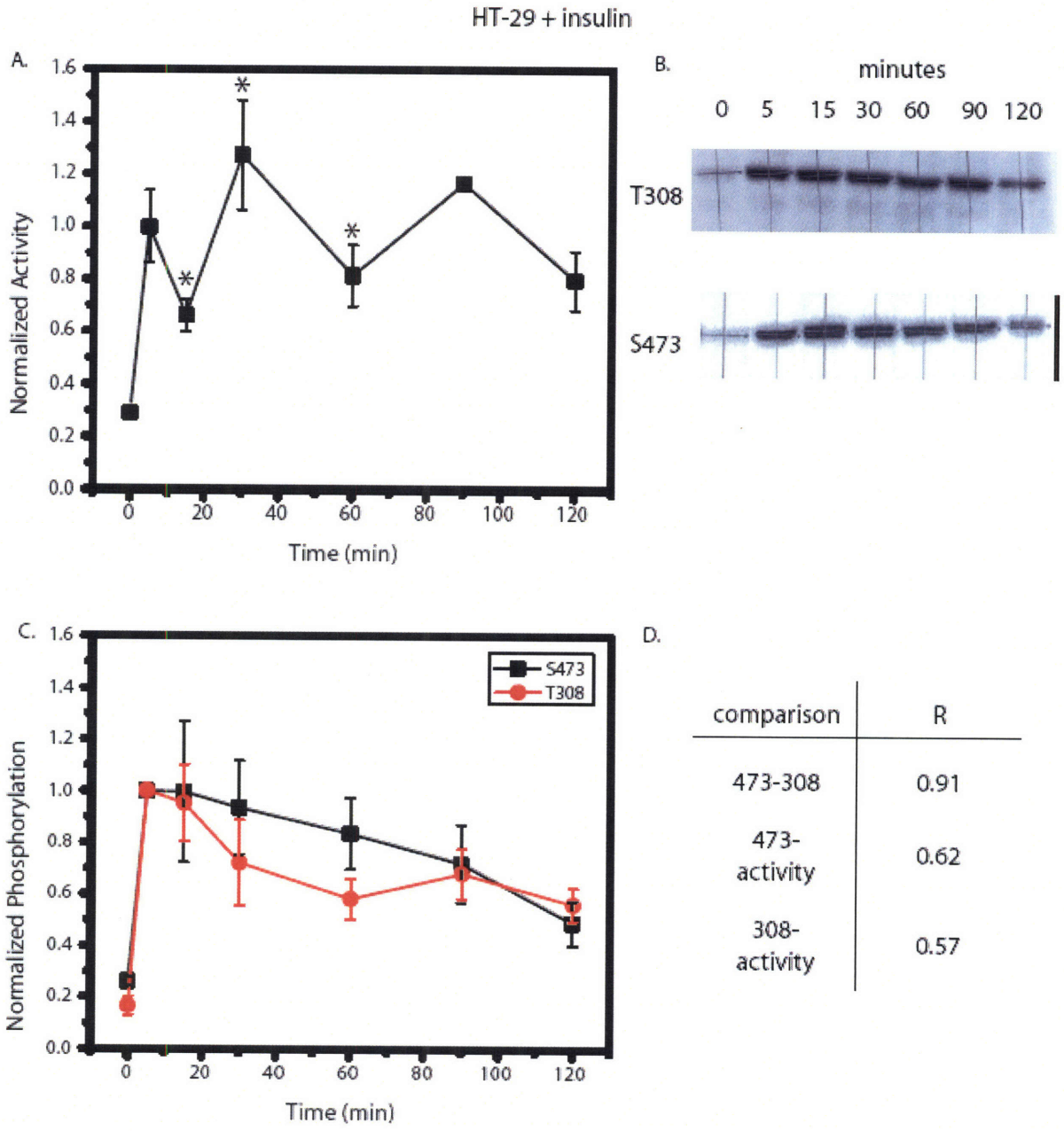


Figure 7-7: HT-29 cells treated with insulin exhibit oscillatory Akt activation and sustained phosphorylation. An *in vitro* kinase activity assay was used to measure Akt activity in HT-29 cells treated with insulin (500 ng/ml) at 0, 10, 15, 30, 60, 90, and 120 minutes (A). Phosphorylation at T308 and S473 was also measured under these conditions using western blot analysis. Shown in (B) are representative blots for T308 and S473 from the three biological replicates measured. Densitometry was used to quantify the net band intensity for all western blots (C). Calculation of the Pearson's correlation between phosphorylation of T308, S473, and kinase activity is shown in (D). All points in the time courses are the average of three biological replicates ± SEM. Time points were normalized to 5 minute kinase activity or phosphorylation levels. * indicates

that the difference between the time point and the one previous to it is significant ($P < 0.05$) at a 95% confidence interval.

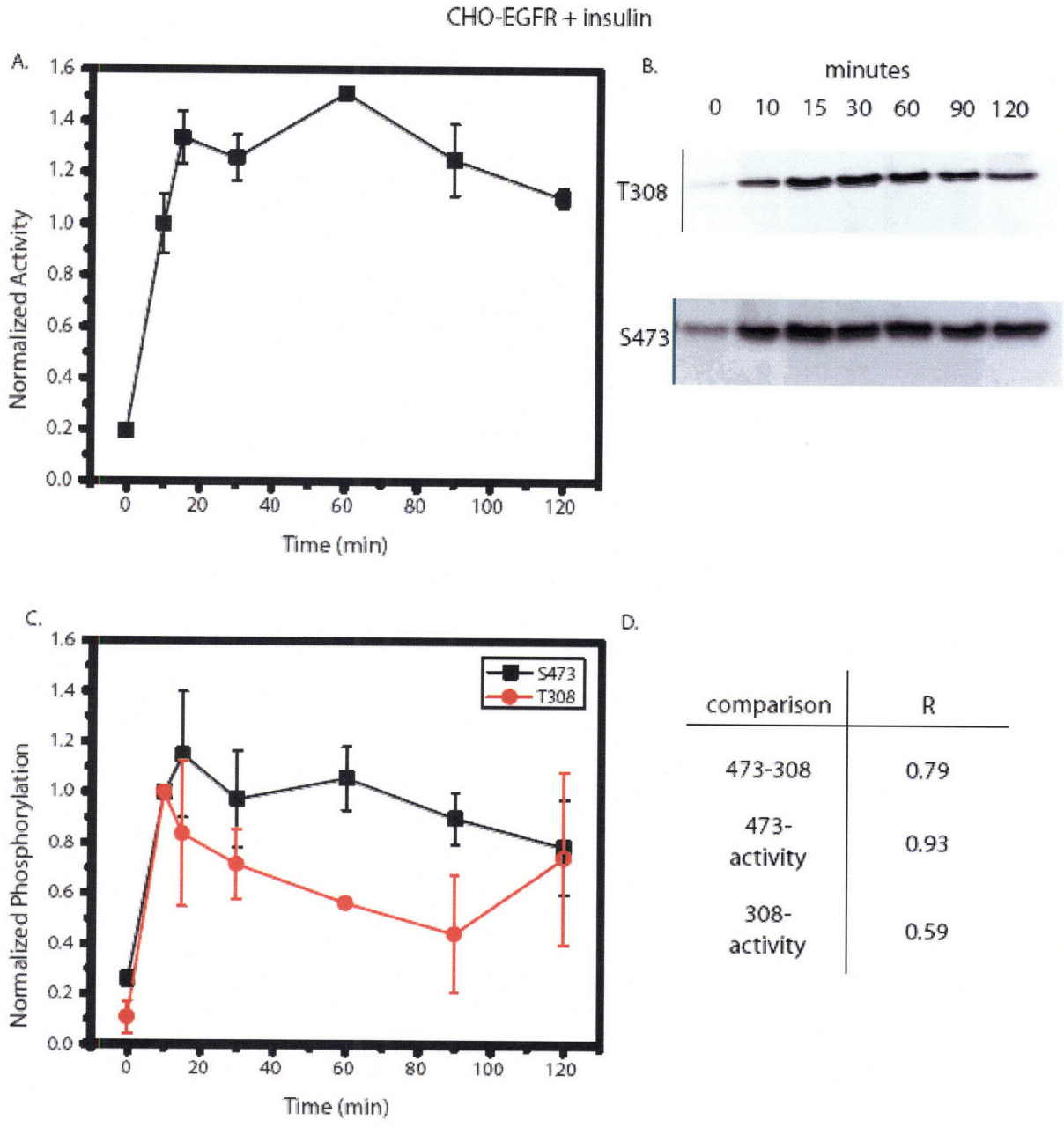


Figure 7-8: CHO-EGFR cells treated with insulin exhibit sustained Akt activation mirrored by S473 but not T308 phosphorylation. An *in vitro* kinase activity assay was used to measure Akt activity in CHO-EGFR cells treated with insulin (500 ng/ml) at 0, 10, 15, 30, 60, 90, and 120 minutes (A). Phosphorylation at T308 and S473 was also

measured under these conditions using western blot analysis. Shown in (B) are representative blots for T308 and S473 from the three biological replicate measured. Densitometry was used to quantify the net band intensity for all western blots (C). Calculation of the Pearson's correlation between phosphorylation of T308, S473, and kinase activity is shown in (D). All points in the time courses are the average of three biological replicates ± SEM. Time points were normalized to 10 minute kinase activity or phosphorylation levels.

7.4 Discussion and conclusions

The relationship between the phosphorylation state of Akt and its catalytic activity is not precisely understood. In this work, we endeavored to understand the accuracy associated with the most common tools used to interrogate Akt function, namely the phospho-specific antibodies against S473 and T308. To do this, we compiled quantitative phosphorylation and kinase activity data spanning two cell lines and as many ligand treatments for seven time points over the course of two hours. As far as we know, this constitutes a uniquely information-rich set of data in which multiple facets of Akt's activation mechanism can be explored in the context of biologically relevant cell lines and ligand treatments.

Since, as mentioned in the introductory section, novel activity-regulating tyrosine phosphorylation sites have been identified in cells stimulated with EGF, we hypothesized that quantitative correlation between T308 or S473 phosphorylation and kinase activity may not be accurate. For cells stimulated with insulin, however, we expected that the correlation would be more accurate, given the fact that much of the seminal work linking phosphorylation and kinase activity was done in the presence of insulin or IGF-1. Counterintuitively, we observed a tight correlation between T308 and

S473 phosphorylation and kinase activity in response to EGF, but under insulin treatment the correlation was weaker, with phosphorylation trends generally failing to match long term kinase activity profiles, particularly in the case of the HT-29 cells where oscillations were observed throughout much of the two hour time course. These data suggest that Akt activity in response to insulin treatment may be regulated through other phosphorylation sites or perhaps through protein-protein interactions that were not disrupted by the wash steps in the *in vitro* kinase assays. Candidate phosphorylation sites might include the afore mentioned tyrosine phosphorylation sites or the autophosphorylated residues T72 and S246, which have been shown to regulate kinase activity in response to insulin stimulation [14]. The data under EGF treatment suggests that although tyrosine phosphorylation is necessary for kinase activity, it may be tightly coupled to S473 and T308 phosphorylation, at least at the level of temporal resolution presented in this study. It is worth noting that under all treatment conditions measurements of either S473 or T308 captured the qualitative activation of Akt kinase such that S473 or T308 phosphorylation measurement may be sufficient depending on the quantitative accuracy one needs for biological interpretation. Finally, in three of the four cellular conditions studied, both T308 and S473 phosphorylation levels correlated strongly, suggesting that measurement of only one of the sites is necessary.

We noted a strong correlation between T308 and S473 phosphorylation during the initial onset of kinase activity. This result was expected, as it is generally accepted that the initial phosphorylation of these two sites is strongly coupled. Nevertheless, studies have also indicated that S473 phosphorylation precedes and promotes T308 phosphorylation to achieve kinase activation [9]. We do not see any evidence of

differential S473 versus T308 early phase phosphorylation, but this may be due to a lack in temporal resolution at early times in our experiment. Although the phosphorylation of T308 and S473 is expected to be tightly coupled, several studies have shown that the dephosphorylation of these two sites is uncoupled [10, 11]. In particular, prior studies using the shellfish toxin okadaic acid suggest that phosphatase activity at the T308 site is not connected to dephosphorylation at the S473 site, with further work suggesting that the phosphatase PP2A is responsible for T308 dephosphorylation [10, 11, 17]. Gao et al. recently showed that a novel phosphatase, PHLPP, is responsible for dephosphorylating the S473 site [10]. Interestingly, the gene for PHLPP is found near a commonly mutated chromosomal region in colon cancers, and Gao et al. showed that the HT-29 colon carcinoma cell line had decreased expression of PHLPP. Thus, we hypothesized that the dephosphorylation of T308 and S473 might be decoupled in HT-29 cells. However, under EGF stimulation, where we observed a transient spike of Akt activity and phosphorylation levels, the rapid dephosphorylation of both S473 and T308 is tightly coupled (Figure 7-5). In the case of insulin treatment, where we observed sustained kinase activity with only partial dephosphorylation over two hours, we again observed a high correlation between T308 and S473 phosphorylation levels (R = 0.91), although on average T308 was dephosphorylated more from 15 to 30 minutes (Figure 7-7). Our results in EGF treated HT-29 cells, where dephosphorylation was significant, indicate that the coordinated dephosphorylation of both T308 and S473 can occur in cell systems where individual phosphatase levels are abnormal. Interestingly, CHO-EGFR cells stimulated with insulin do show evidence of decoupled dephosphorylation, where T308 levels decline more

rapidly than both S473 or activity levels (Figure 7-8). This finding is consistent with Yamada et al.'s observations in insulin treated CHO cells, where they observed rapid T308 dephosphorylation not reflected in either activity or S473 levels.

In sum, our data help to delineate the confidence with which researchers can use commercially available phospho-specific antibodies to understand signaling downstream of the Akt kinase. In addition, the quantitative approach taken allows for a greater understanding of the coordinate regulation of S473 and T308 phosphorylation levels. Future work focused on the measurement of more phosphorylation sites in the case of insulin treatment and with more accurate measurement technologies should enable further insights from this type of experimental approach.

References

1. Brazil, D.P., Z.Z. Yang, and B.A. Hemmings, *Advances in protein kinase B signalling: AKTion on multiple fronts.* Trends Biochem Sci, 2004. **29**(5): p. 233-42.
2. Hanada, M., J. Feng, and B.A. Hemmings, *Structure, regulation and function of PKB/AKT--a major therapeutic target.* Biochim Biophys Acta, 2004. **1697**(1-2): p. 3-16.
3. Vivanco, I. and C.L. Sawyers, *The phosphatidylinositol 3-Kinase AKT pathway in human cancer.* Nat Rev Cancer, 2002. **2**(7): p. 489-501.
4. Feng, J., et al., *Identification of a PKB/Akt hydrophobic motif Ser-473 kinase as DNA-dependent protein kinase.* J Biol Chem, 2004. **279**(39): p. 41189-96.

5. Persad, S., et al., *Regulation of protein kinase B/Akt-serine 473 phosphorylation by integrin-linked kinase: critical roles for kinase activity and amino acids arginine 211 and serine 343.* J Biol Chem, 2001. **276**(29): p. 27462-9.

6. Toker, A. and A.C. Newton, *Akt/protein kinase B is regulated by autophosphorylation at the hypothetical PDK-2 site.* J Biol Chem, 2000. **275**(12): p. 8271-4.

7. Sarbassov, D.D., et al., *Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex.* Science, 2005. **307**(5712): p. 1098-101.

8. Alessi, D.R., et al., *Mechanism of activation of protein kinase B by insulin and IGF-1.* Embo J, 1996. **15**(23): p. 6541-51.

9. Scheid, M.P., P.A. Marignani, and J.R. Woodgett, *Multiple phosphoinositide 3-kinase-dependent steps in activation of protein kinase B.* Mol Cell Biol, 2002. **22**(17): p. 6247-60.

10. Gao, T., F. Furnari, and A.C. Newton, *PHLPP: a phosphatase that directly dephosphorylates Akt, promotes apoptosis, and suppresses tumor growth.* Mol Cell, 2005. **18**(1): p. 13-24.

11. Yamada, T., et al., *3-phosphoinositide-dependent protein kinase 1, an Akt1 kinase, is involved in dephosphorylation of Thr-308 of Akt1 in Chinese hamster ovary cells.* J Biol Chem, 2001. **276**(7): p. 5339-45.

12. Chen, R., et al., *Regulation of Akt/PKB activation by tyrosine phosphorylation.* J Biol Chem, 2001. **276**(34): p. 31858-62.

13. Conus, N.M., et al., *Direct identification of tyrosine 474 as a regulatory phosphorylation site for the Akt protein kinase.* J Biol Chem, 2002. **277**(41): p. 38021-8.

14. Li, X., et al., *Autophosphorylation of Akt at threonine 72 and serine 246. A potential mechanism of regulation of Akt kinase activity.* J Biol Chem, 2006. **281**(19): p. 13837-43.

15. 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.

16. Janes, K.A., et al., *A high-throughput quantitative multiplex kinase assay for monitoring information flow in signaling networks: application to sepsis-apoptosis.* Mol Cell Proteomics, 2003. **2**(7): p. 463-73.

17. Andjelkovic, M., et al., *Activation and phosphorylation of a pleckstrin homology domain containing protein kinase (RAC-PK/PKB) promoted by serum and protein phosphatase inhibitors.* Proc Natl Acad Sci U S A, 1996. **93**(12): p. 5699-704.

Chapter 8 Applications and concluding remarks

8.1 Applying computational modeling to drug discovery and development

8.1.1 Introduction and Motivation for Use of Models

Prediction is the attempt to use existing knowledge to foretell an event before it happens. Whether it is the biologist trying to predict how target inhibition will affect cell behavior, the physician trying to predict how a drug will affect a patient, or the manager trying to predict future return on investment, prediction plays a vital role in the pharmaceutical industry. Breakthroughs in the fields of genetics, biochemistry and molecular biology have increased our ability to understand and predict behavior in biological systems. Here we argue that computational modeling based on biological information can be used to extend the limits of our understanding toward predictive accuracy. Models used to simulate cellular or human biology produce reliable data, new hypotheses and can translate information between *in vitro* screens, cell-based assays and ultimately patients. This extension of knowledge is valuable to the pharmaceutical industry for novel product generation. Even in the absence of novel products, incorporation of computational modeling in the ways outlined below could save millions of dollars based on increased efficiency.[1, 2]

A coarse-grained schematic of the pharmaceutical research and development (R&D) pipeline is shown in Figure 8-1. We have identified three areas where computational modeling has potential to substantially impact efficiency and development. The first area is cell-signal behavior, where the application of models characterizes how lead compounds affect intracellular signaling. The second area is signal-response behavior, where models predict cellular phenotype from signaling information. The third area is physiology, in which models are used to simulate clinical outcomes. Additionally, each class of model can help identify new drug targets. We address each application area separately, highlighting important work relevant to the pharmaceutical industry.

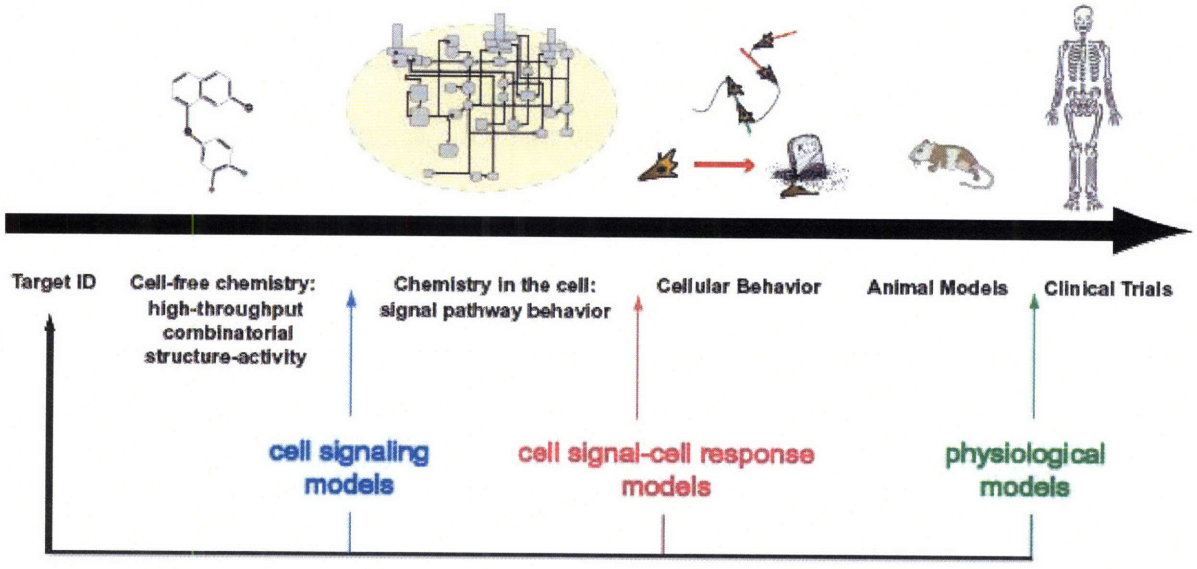


Figure 8-1: Areas of impact for computational modeling in the pharmaceutical R&D process. A course-grained diagram of the R&D process illustrates three potential areas for model application. Cell signaling models simulate intracellular signaling dynamics and predict drug effects on signaling. Cell signal-cell response models correlate intracellular signals to cell behaviors such as migration and apoptosis. These models predict drug effects on cell behavior.

In addition to specific applications, there is also a natural role for modeling to link traditional biology and high-throughput informatics analysis (Figure 8-2). For instance, the construction of a signaling model begins with an assembly of molecular interactions, rate parameters and spatial restrictions. Informatics groups analyze high-throughput datasets (*i.e.* gene-chip arrays, gene sequencing results, mass spectrometry results, yeast two-hybrid results), using methods such as clustering or spacing alignments, and integrate results with data from other in-house biological experiments and from literature (obtained *via* text mining). The data are then further organized into ontologies [3]. A model is constructed from a subset of these data and is then validated using traditional biology experiments. If the model captures experimental trends, it is used to generate predictions or hypotheses that suggest new biological experiments. The results of these experiments either further validate the model or identify novel biology that is then incorporated into the model. This interplay between informatics, modeling and traditional biology enables the focused use of large datasets to solve biologically relevant problems.

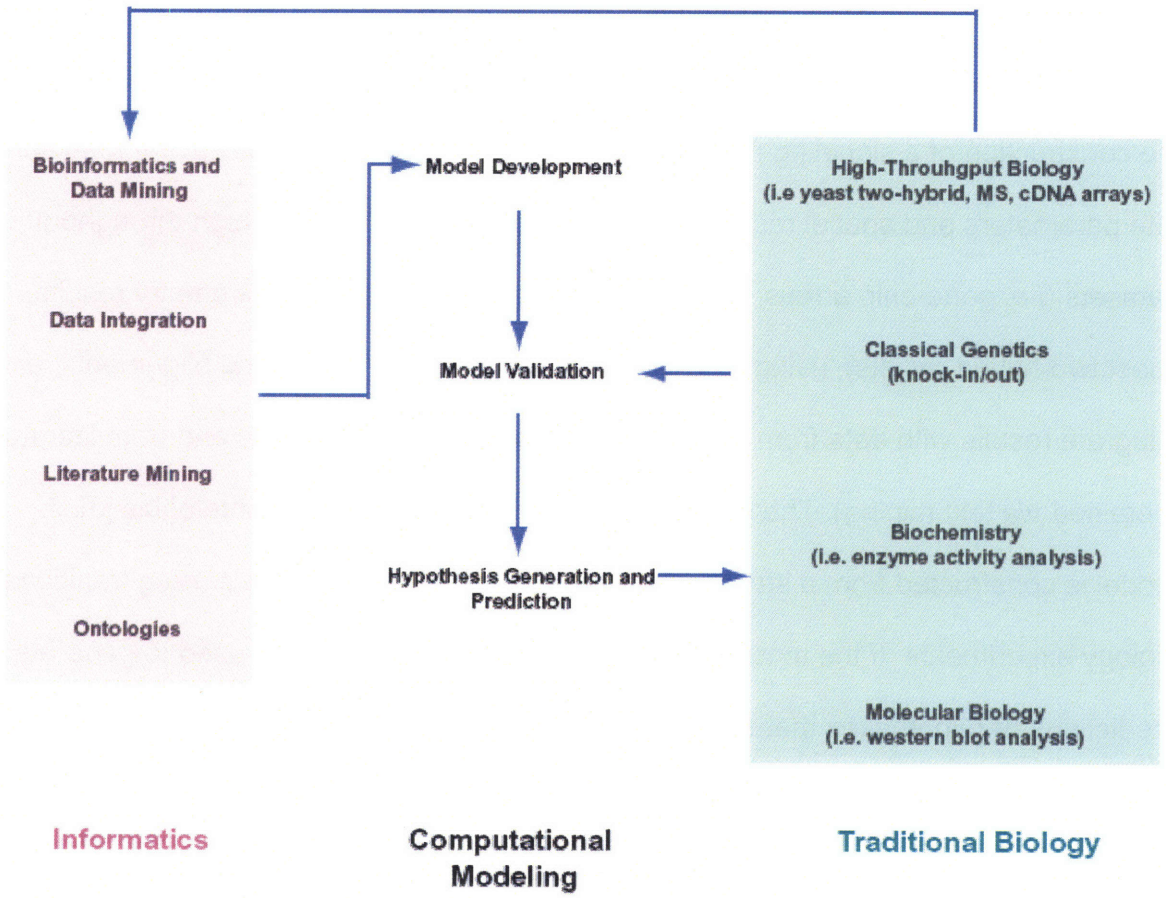


Figure 8-2: Computational modeling in the R&D workflow. Computer models rely on data from both informatics and traditional biology applications. As such, they help coordinate the use of informatics to answer biological questions. Informatics groups analyze large biological datasets, mine scientific literature for data, integrate disparate forms of data, and organize databases for further use. Computational models are developed using subsets of stored data. Models are validated against results from genetic, biochemical, or molecular biology experiments. After validation, models are used to predict and interpret novel biology. Resulting hypotheses are tested using traditional biology techniques. If predicted data do not correspond to experimental measurements, the model is altered to incorporate the new biological information.

We have restricted the scope of our discussion to selected modeling efforts with emerging relevance to the pharmaceutical industry. We do not, however, address many models employed successfully by the pharmaceutical industry today. For instance,

pharmacokinetic (PK) models, which we do not cover here, are perhaps the most significant class of models being used today. In addition, we do not discuss myriad other metabolic models that have been successfully employed at cellular and animal levels. Nevertheless, we hope to demonstrate how new advances in computational modeling can integrate into the R&D workflow, and in so doing we attempt to establish a framework for thinking about the application of diverse types of models to the pharmaceutical industry.

8.1.2 Cell Signaling Models

Defects in signal transduction underlie many diseases of interest to pharmaceutical companies. For instance, dysregulation of conserved protein tyrosine kinase pathways leads to a variety of cancers [4]. Individual signaling proteins inside the cell are often the target of small-molecule drugs, whereas many antibody drugs target the receptors controlling signaling cascades. Here, we highlight a group of models that have been used to describe, computationally, signaling pathways relevant to disease. These models, in comparison to the signal-response models discussed in the next section, are more highly specified in their molecular details.[5] Typically, ordinary differential equations (ODEs) are used to describe mass-action kinetics and system behavior. Experimental measurement of reaction rates, concentrations, molecular interactions, and trafficking parameters are essential for the construction of such models. The level of detail necessary varies from system to system, but many signal-

transduction pathways can be modeled using a combination of measured values, fitted parameters, and coarse-grained descriptions of interactions.[6]

Models that describe signaling pathways are important in pharmaceutical research for three main reasons: i) they often capture non-intuitive signal behavior and identify novel molecular function; ii) they allow researchers to experiment *in silico* across a wide range of conditions (*i.e.* receptor numbers, ligand concentrations, phosphorylation rates), thus saving experimental resources and identifying important further experiments and iii) they serve as a database for much of the known information about a particular pathway.

Lee *et al.* used ODE-based models together with experiment to analyze the Wnt signaling pathway, a pathway critical in disease physiology such as cancer[7]. Model construction from published data revealed that further experiments to measure total concentrations and rates of dissociation were necessary. Successful implementation of the model showed that adenomatous polyposis coli (APC) and axin, both of which coordinate the degradation of the Wnt pathway effector β-catenin, bind proteins in different ways, with axin binding randomly and APC binding in an ordered manner to promote degradation. The model also highlighted APC-dependant axin degradation, a mechanism potentially important in explaining β-catenin levels in response to APC mutation. Using the model, the effects of perturbing seven different Wnt pathway molecules were explained, exemplifying the potential target selection and drug testing applications of such an approach. In addition, the relevant investment compared with exclusive experimental testing is considerably smaller. Model predictions also indicate that slight changes in APC function lead to significant buildup of β-catenin levels. In

colorectal cancer, APC mutations give rise to elevated β-catenin levels, resulting in a cancerous phenotype. Potential therapies based upon these model predictions might include drugs that inhibit protein phosphatase 2A (PP2A) or breakdown T-cell factor. The model indicates that PP2A inhibition would be more effective in terms of β-catenin degradation, but this would also affect axin levels to a greater extent, which might in turn perturb other signaling pathways reliant on axin.[7, 8]

Hendriks *et al.* developed a detailed model of ErbB receptor signaling dynamics coupled to a receptor trafficking model [9]. ErbB receptors and their signaling pathways are implicated in a variety of cancers and are the target of many therapeutic compounds. [4] The model was developed using binding, dimerization and trafficking rate constants from literature. Hendriks *et al.* simulated time-resolved phosphorylation profiles for 3 types of ErbB receptors under more than 10 ligand stimulating conditions. These simulations required less than 24 hours on a desktop computer; by comparison, the equivalent experimental data would have required weeks of effort and resources at the bench. Further simulation produced phosphorylation profiles under varying assumptions, including dephosphorylation of receptors at the surface versus exclusive dephosphorylation of receptors internalized in the endosome. Comparison with experimental data revealed that ErbB receptor dephosphorylation occurred primarily in the endosomal compartment. While many drugs target the extracellular domain of ErbB receptors to prevent ligand binding and subsequent phosphorylation, the model predicts that shunting more receptors per unit time to the endosome will attenuate phophorylation. Inhibition of Sprouty, a class of molecules that are known to inhibit receptor internalization through interaction with Casitas B-lineage (Cbl), could increase

trafficking to the endosome, and thus provides a molecular target for testing this model-derived hypothesis.[10] In addition, scientists could use the model to predict the degree of elevated trafficking needed to achieve a given reduction in ErbB phosphorylation levels. But how much reduction in phosphorylation is enough to affect downstream signaling? Again, computational modeling can help. Well known downstream effectors of the ErbB pathways include extracellular-regulated kinase (ERK) and protein kinase B (PKB or AKT). Hatakeyama *et al.* and Schoeberl *et al.* have constructed models that simulate the effect of receptor phosphorylation on these important signaling intermediates. The models identify how variation in total amounts of phosphorylation and rates of phosphorylation affect downstream signaling. Importantly, they also reveal how signaling molecules in distinct-but-connected pathways are regulated after ligand activation or signal inhibition.[11, 12]

All of the above efforts were effective because they were validated using experimental methods, they explored many more hypotheses than would be experimentally feasible and they highlighted non-intuitive, but important, regulatory schemes for signaling. Recent documented efforts indicate that cell signaling models are being successfully employed by pharmaceutical companies. For instance, AstraZeneca recently used a computational approach to link the efficacy of the cancer drug Iressa to impaired receptor internalization and reliance on downstream AKT signaling.[13] The further application of these types of models relies on clear relationships between intracellular signaling proteins and cell behavior or higher-level processes. The general approach suffers when these relationships are poorly defined. The next class of models surmounts this difficulty to predict complex cellular behavior.

8.1.3 Signal-Response Models

The cells that make up the human body engage in a large variety of behaviors: cell migration, differentiation, proliferation, and programmed death are just a few of the many functions cells carry out. Interestingly, it has been hypothesized that no more than 20 signal transduction cascades control the seemingly endless list of cell behaviors observed in humans[7]. How does the cell, then, use these 20 or so cascades to coordinate cell behavior? Part of the answer is that each pathway can be activated in quantitatively different ways. For instance, in PC12 neuronal cells, transient activation of the mitogen-activated protein kinase (MAPK) pathway leads to proliferation whereas sustained activation of this exact same pathway leads to differentiation.[14] Another part of the answer lies in the fact that multiple pathways can be used to control one behavior. For instance, hepatocyte growth factor (HGF)-stimulated neuronal migration is regulated by both the MAPK and phosphatidylinositol 3-kinase (PI3K) signaling pathways [15]. Both of these facts suggest that to correct aberrant cellular behavior with drugs requires quantitative knowledge about multiple signaling proteins (that is, multivariate datasets). Multivariate datasets can then be used to understand cellular decision-making processes in the context of computational models.

The signaling pathway models discussed earlier give us insight into how the MAPK pathway might be manipulated to yield either transient or sustained ERK activation. Mapping of this activation, however, onto cellular behavior is an arguably more difficult task. Whereas ODE level models are becoming more prevalent for

describing signaling pathways, there are very few models that can accurately connect signaling pathways to cellular behavior at this level of mathematical description. The problem, therefore, requires the use of more abstracted signaling models[5]. Abstracted models identify statistical relationships between signals and behavior, which suggest causal signal-behavior relationships that can be further probed using molecular biology or genetic approaches. Here, we address a few recent models that have used multivariate cell signaling data to reveal governing principles of cellular behavior. Predictions made on the basis of these models can reveal how a drug, or class of drugs, will affect a given cellular behavior.

Many recent efforts have focused on using modeling to identify genomic and proteomic groups of molecules responsible for disease-relevant cellular behaviors. For instance, Zheng *et al.* used cDNA microarrays together with 2D gel electrophoresis and mass spectrometry to study the molecular effects of an acute promyelocytic leukemia (APL) cell line treated with retinoic acid (RA) and arsenic trioxide (ATO) [16]. The question being asked was: how do downstream signaling events coordinate a known program of differentiation and apoptosis? The answer to this question may shed some light on the signaling events responsible for clinical efficacy in APL patients co-treated with RA/ATO in conjunction with chemotherapy. Zheng *et al.* used a computational technique called self organizing maps (SOMs) to cluster signaling data and then characterized sets of signals important for the differentiation-*vs*-apoptotic cell program. They found that, among other things, activation of the interferon (IFN) and calcium signaling pathways coordinated differentiation and apoptosis. They also identified a number of transcription factors important for coordinated cell behavior. By isolating sets

of genomic and proteomic changes associated with a specific therapeutic result, the authors have identified drug mechanism and revealed important sets of pathways for future drug development. Conceptually similar studies have measured quantitative proteomic and genomic data, and then used computationally-aided organization of this data to interpret the role of signaling groups on cellular response [17, 18]. Recent efforts establishing novel ways to integrate multiple types of measurement datasets should enable future successes for the application of this approach [19, 20].

Although the previously-mentioned signal-response studies rely on computation to identify co-regulation of signals, their goal is generally one of classification (also referred to as profiling or signature analysis) rather than prediction. To do this, quantitative measurement of the signal/transcription state and cellular response is required. We now discuss recent efforts that have established computational models to understand cellular behavior in the context of cell signaling. Models of this type are very useful for the pharmaceutical industry, as they allow scientists to alter signals and then predict how cellular behavior changes *in silico*. Janes *et al.* describe a procedure based on linear modeling (partial least squares regression), whereby approximately 8000 intracellular signals were correlated with more than 1000 apoptosis-related cellular responses [21]. The model computationally organizes the vast amount of signaling data measured, just as in previously mentioned studies, but takes the analysis a step further by deriving a set of parameters that map the signaling values onto the apoptosis measurements. Importantly, Janes *et al.* [21] tested their model by perturbing the cellular signal state and then comparing experimentally-measured apoptotic values with those predicted from the model. For instance, C225, an antibody raised against

epidermal growth factor receptor (EGFR), was used to treat a human colon carcinoma cell line. Excellent agreement between predicted and measured levels of apoptosis validated the model and suggests that it can be used to understand apoptotic response under a wide range of signal perturbations (induced by drugs or cytokines). Indeed, the humanized form of C225, known as Cetuximab, was approved by the FDA for treatment of advanced colorectal cancer, demonstrating that these approaches have direct pharmaceutical relevance. Furthermore, a recent published effort between Pfizer and academic researchers shows that signal-response modeling is already being applied successfully within the context of the pharmaceutical industry [22]. In this study, Clayton *et al.* created a linear model, using NMR-measured pre-dose metabolite profiles, to predict the effect of paracetamol (acetaminophen) on rat excretion profiles and liver damage [22]. In addition, we recently applied this modeling technique to understand the effects of ErbB2 receptor over-expression (implicated in a large number of breast cancers) on cell proliferation and migration in human breast epithelial cells (unpublished data), suggesting that this technique is broadly applicable to disease-relevant cell systems.

Sometimes, for systems that have been studied in great detail, it becomes possible to create a more mechanistic model of cell behavior based on intracellular signals. Gene Network Sciences (GNS, Ithaca, NY, USA) has developed models for human cell proliferation and apoptosis to study potential anti-cancer strategies. Using a combination of known protein interactions and network inference, Christopher *et al.* describe a model based largely on differential equations that predicts cell proliferation [23]. While highly promising, it is worth noting that deterministic modeling in the area of

cell behavior has proven very difficult, in part because of the level of detail required. This fact prompted GNS to develop abstracted inference models to connect their deterministic models to cell behavior in the absence of detailed signal-pathway information [24].

8.1.4 Physiological Models

The previous section dealt with models that reveal relationships between cell signaling and function. Given the current state of knowledge, these approaches face great challenges in translating their predictions to clinically measurable outcomes. Although breakthroughs in the fields of tissue engineering and the use of more physiologically relevant cell systems have aided in addressing this issue, the problem of understanding physiology through the use of cell-based computational models remains a difficult one. For models predictive of physiology to be feasible, a system has to have been extensively studied or describable at a high level of abstraction. Given the small number of systems that presently fit these criteria, efforts in this field have produced considerably fewer results than the approaches mentioned earlier in this review. There are, however, a handful of encouraging results that have been published and an increasing number of labs and companies are becoming involved in such efforts.

Noble describes a remarkable computational model of the heart that provides a unified description of organ level physiology in terms of protein level biology. The model provides non-intuitive explanations for how anti-arrhythmia drugs might work. Extensive knowledge of signaling pathways, cell-cell organization, and the tissue geometry of the

heart made this project possible [25]. Related efforts, such as the Physiome Project, attempt to use computation to describe many more systems from the protein level through to the organ level [26]. Other groups have tried to model pathophysiology starting with an organ level (or sometimes even more abstracted) model and then adding layers of information as necessary [27]. For instance, mathematical models of Type 2 Diabetes have been used for over a decade to understand key parameters pertaining to pathology in patients, such as insulin sensitivity and beta-cell function [28]. Entelos (Foster City, CA, USA) has utilized a 'top-down' modeling approach to generate ODE-based models at the organ level. Lewis *et al.* used this framework to explore the pathophysiology of asthma and found that the resulting model could capture the acute and chronic characteristics of asthma. Importantly, their model also correctly predicted a lack of clinical inefficacy for humanized anti-interleukin-5 (IL-5) antibodies[27, 29].

Although the number of models that allow for this type of powerful prediction are few, it is clear from studies such as Stokes *et al.* that the potential benefits for a pharmaceutical company are extremely high. These studies can be used to not only to predict the clinical outcome of a particular drug, but also to identify novel interventions that can 'front load' the R&D pipeline with physiologically relevant targets.

8.1.5 Conclusions and Future Directions in Industry

Computational models address a key issue in the pharmaceutical industry: prediction. Whereas many traditional biological studies present insightful descriptions, the application of the hypotheses to new parameter space such as different ligand

concentrations or receptor numbers is often difficult. Computational modeling solves this problem by bringing existing knowledge together in a quantitative framework that allows scientists to predict the effects of system perturbations. Of course, models do not always predict correctly. When models fail, the underlying assumptions are flawed and fixing this depends upon identifying areas that require further experimentation. Companies best able to execute tight integration between modeling and experiment together with repeated iteration of the modeling/experiment cycle will reap the greatest benefit from the described computational approaches.

Computational models fit within the workflow of the pharmaceutical R&D pipeline, serving to coordinate and explain information being generated in both biological and informatics groups. How well these models serve their function will depend on the effective training of scientists that possess both biological intuition and computational skills. The *in silico* component in research must still be coupled with hypothesis-driven experimental design and is not a substitute for the more important *in cerebro* component. The interrogation of models by those who have biological understanding will be vital for the development of successful models. Most large pharmaceutical firms currently have computational modeling groups, including Johnson & Johnson, Eli Lilly, AstraZeneca, Pfizer, Novartis, GlaxoSmithKline, and Merck. In addition, a number of small model-focused companies such as Gene Network Sciences, Entelos, BG Medicine, BioSeek, and Merrimack Pharmaceutical are developing state-of-the-art computational models for the pharmaceutical industry [30]. We believe that the most successful models will not only provide predictive power but will also be scalable, meaning that models currently appropriate for different phases in the R&D pipeline

should be mutually compatible in anticipation of information that will connect disparate R&D stages. Specifically, since PK models have already proven essential for the drug development process, new models capable of integrating with PK models will be most useful for pharmaceutical companies. As new computational tools become available, companies successful in applying modeling will have a competitive advantage derived from increases in predictive power along the R&D pipeline.

8.2 Concluding remarks

It is my hope that the work presented in this thesis will serve as a template for future investigators interested in exploring cell signaling and cell behavior. In Chapters 2 and 3, I introduced new technologies for the measurement of kinase phosphorylation and cell migration. It is hoped that these technologies will be further applied to systems of interest, and in particular, that the migration assay will be used for high-throughput applications such as drug or siRNA screens. In addition, it is hoped that the work in Chapter 2 prompts future computational and experimental studies into the importance of directionally persistent migration. In Chapters 4 and 5, I presented the application of PLSR models to understand the effects of HER2 overexpression on cell signaling and cell behavior. It is hoped that the many hypotheses generated by the models will be further tested using the tools of molecular biology. It is further hoped that investigators interested in constructing signal-response models will take advantage of the PLSR-based advances documented in Chapter 5. Chapter 6 explored the role of signal state context and off-target effects due to network crosstalk in the efficacy of drugs. It is hoped that this work increases the attention paid to the network-signaling properties of

drugs and that the modeling approaches used to understand these properties will be applied to understand the efficacy of future drugs. Finally, a quantitative approach to understanding Akt phosphorylation and activity was presented in Chapter 7. It is hoped that this work will not only prompt future inquiry into the regulatory mechanisms of Akt, but also serve as a blueprint for further work targeting other important kinases. In sum, I hope that this work plays a small part in helping to develop a more predictive theory of biological signaling and the behavior it governs.

References

1. Butcher, E.C., *Can cell systems biology rescue drug discovery?* Nat Rev Drug Discov, 2005. **4**(6): p. 461-7.
2. DiMasi, J.A., R.W. Hansen, and H.G. Grabowski, *The price of innovation: new estimates of drug development costs.* J Health Econ, 2003. **22**(2): p. 151-85.
3. Searls, D.B., *Data integration: challenges for drug discovery.* Nat Rev Drug Discov, 2005. **4**(1): p. 45-58.
4. Blume-Jensen, P. and T. Hunter, *Oncogenic kinase signalling.* Nature, 2001. **411**(6835): p. 355-65.
5. 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.
6. Kholodenko, B.N., *Cell-signalling dynamics in time and space.* Nat Rev Mol Cell Biol, 2006. **7**(3): p. 165-76.
7. Lee, E., et al., *The roles of APC and Axin derived from experimental and theoretical analysis of the Wnt pathway.* PLoS Biol, 2003. **1**(1): p. E10.

8. Behrens, J., *The role of the Wnt signalling pathway in colorectal tumorigenesis.* Biochem Soc Trans, 2005. **33**(Pt 4): p. 672-5.

9. B.S. Hendriks, J.C., J.M. Burke, J.M. Beusmans, D.A. Lauffenburger and D. de Graaf, *Computational modeling of ErbB family phosphorylation dynamics in reponse to transforming growth factor alpha and heregulin indicates spatial compartmentation of phosphatase activity.* IEE Proc.-Syst. Biol., 2006. **153**(1): p. 22-33.

10. Kim, H.J. and D. Bar-Sagi, *Modulation of signalling by Sprouty: a developing story.* Nat Rev Mol Cell Biol, 2004. **5**(6): p. 441-50.

11. Hatakeyama, M., et al., *A computational model on the modulation of mitogen-activated protein kinase (MAPK) and Akt pathways in heregulin-induced ErbB signalling.* Biochem J, 2003. **373**(Pt 2): p. 451-63.

12. Schoeberl, B., et al., *Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors.* Nat Biotechnol, 2002. **20**(4): p. 370-5.

13. B.S. Hendriks, G.J.G., R. Benson, D. Kenyon, M. Lazzara, J. Swinton, S. Beck, M. Hickinson, J.M. Beusmans, D.A. Lauffenburger and D. de Graaf, *Decreased internalization of ERBB1 mutants in lung cancer is linked with a mechanism conferring sensitivity to gefitinib.* IEE Proc.-Syst. Biol., 2006. **in press**.

14. Marshall, C.J., *Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation.* Cell, 1995. **80**(2): p. 179-85.

15. Segarra, J., et al., *Combined Signaling through ERK, PI3K/AKT, and RAC1/p38 Is Required for Met-triggered Cortical Neuron Migration.* J Biol Chem, 2006. **281**(8): p. 4771-8.

16. Zheng, P.Z., et al., *Systems analysis of transcriptome and proteome in retinoic acid/arsenic trioxide-induced cell differentiation/apoptosis of promyelocytic leukemia.* Proc Natl Acad Sci U S A, 2005. **102**(21): p. 7653-8.

17. Hirai, M.Y., et al., *Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in Arabidopsis thaliana.* Proc Natl Acad Sci U S A, 2004. **101**(27): p. 10205-10.

18. Kao, K.C., et al., *Transcriptome-based determination of multiple transcription regulator activities in Escherichia coli by using network component analysis.* Proc Natl Acad Sci U S A, 2004. **101**(2): p. 641-6.

19. Hwang, D., et al., *A data integration methodology for systems biology.* Proc Natl Acad Sci U S A, 2005. **102**(48): p. 17296-301.

20. Hwang, D., et al., *A data integration methodology for systems biology: experimental verification.* Proc Natl Acad Sci U S A, 2005. **102**(48): p. 17302-7.

21. 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.

22. Clayton, T.A., et al., *Pharmaco-metabonomic phenotyping and personalized drug treatment.* Nature, 2006. **440**(7087): p. 1073-7.

23. Christopher, R., et al., *Data-driven computer simulation of human cancer cell.* Ann N Y Acad Sci, 2004. **1020**: p. 132-53.

24. Aksenov, S.V., et al., *An integrated approach for inference and mechanistic modeling for advancing drug development.* FEBS Lett, 2005. **579**(8): p. 1878-83.

25. Noble, D., *Systems biology and the heart.* Biosystems, 2006. **83**(2-3): p. 75-80.

26. Hunter, P.J. and T.K. Borg, *Integration from proteins to organs: the Physiome Project.* Nat Rev Mol Cell Biol, 2003. **4**(3): p. 237-43.

27. Butcher, E.C., E.L. Berg, and E.J. Kunkel, *Systems biology in drug discovery.* Nat Biotechnol, 2004. **22**(10): p. 1253-9.

28. Kansal, A.R., *Modeling approaches to type 2 diabetes.* Diabetes Technol Ther, 2004. **6**(1): p. 39-47.

29. Lewis, A., et al., *The roles of cells and mediators in a computer model of chronic asthma.* INTERNATIONAL ARCHIVES OF ALLERGY AND IMMUNOLOGY, 2001. **124**(1-3): p. 282-286.

30. Mack, G.S., *Can complexity be commercialized?* Nat Biotechnol, 2004. **22**(10): p. 1223-9.

Chapter 9 Appendices

Appendix 1

```
% 2D migration analysis by Hyung-Do Kim
% Purpose: Quantitation of cell migration trajectories from coordinate
data

clear;
addpath(pwd,'-begin');
datadir = uigetdir;
cd(datadir);

% Enter experiment name
expname = input('\nEnter the experiment name: \n','s');
mkdir('Figures');
mkdir('Tiffs');

% Input file format: .txt
% Input text file:
filename = input('\nEnter the text file: \n','s');
data = load(filename);

% Enter time interval
interval_i = input('\nEnter the time interval in minutes: \n');
interval = interval_i./60;

% Enter image conversion factor and convert
global conversion
conversion=input('\nEnter the um/pixel conversion factor: \n');
data = data.*conversion;

% Calculate mean square displacement using non-overlapping intervals
msdnonoverlap;
fprintf('\nMSD calculation completed\n');
wait = input('Press Enter to continue','s');

% Fit speed and persistence using generalized nonlinear least-squares
% regression
prwfit;

% Make arrays of data and Save data into csv file.
% First save .mat file with variables
save([expname '.mat'],'n_cells','cell','xx','yy','distance','msd',...

'sigma_msd_cell','int_counter','avg_total_msd','max_msd','interval',...

'Scalc','Sfit','Pcalc','Pfit','Calcresnorm','Fitresnorm','Pmisfit',...
    'speed_mean','speed_median','speed_sem','speed_max','speed_min',...
```

```
'persistence_mean','persistence_median','persistence_sem','persistence_
max',...

'persistence_min','CI_mean','CI_median','CI_sem','CI_max','CI_min',...
    'mu',
'mu_mean','meanpathlength','meanpathlength_mean','total_int');

% CSV file of individual cell data
% Row of parameters
csvname = [expname '_ind.csv'];
parind = ['Cell #,' 'Number of Intervals,' 'Calc. Speed Mean [um/hr],'
'Calc. Speed SEM [um/hr],' 'Fit Persistence [min],'...
    'Fitting Residuals,' 'Chemotactic Index,' 'Total Path Length [um],'
'Total Displacement [um],' ...
    'Random Motility C. [um^2/hr],' 'Mean Path Length [um]'];
fid = fopen(csvname, 'wt');
fprintf(fid, '%s', parind);
fclose('all');
% Data
indM = [[1:n_cells]' total_int' Scalc(:,1) Scalc(:,2) Pcalc(:,1).*60
Calcresnorm(:,1) CI(:,1) ...
    total_cell_path_length' total_cell_displacement' mu(:,1)
meanpathlength(:,1)];
dlmwrite(csvname, indM, '-append','roffset',1);
clear indM;

% CSV file of pooled data set
% Column of parameters
csvname = [expname '_pool.csv'];
parpool = ['Number of cells,','Speed Mean [um/hr],','Speed SEM
[um/hr],',...
    'Speed Median [um/hr],','Speed Max [um/hr],','Speed Min
[um/hr],',...
    'Persistence Mean [min],','Persistence SEM [min],','Persistence
Median [min],',...
    'Persistence Max [min],','Persistence Min [min],','Persistence
Misfit,','CI Mean,','CI SEM,',...
    'CI Median,','CI Max,','CI Min,','Random Motility C
[um^2/hr],','Mean Path Length [um],'];
fid = fopen(csvname, 'wt');
fprintf(fid, '%s', parpool);
fclose('all');
% Data
poolM = [n_cells speed_mean speed_sem speed_median speed_max speed_min
persistence_mean.*60 ...
    persistence_sem.*60 persistence_median.*60 persistence_max.*60
persistence_min.*60 Pmisfit(1) CI_mean CI_sem ...
    CI_median CI_max CI_min mu_mean meanpathlength_mean];
dlmwrite(csvname, poolM, '-append','roffset',1);
clear poolM


% msdnonoverlap.m - Hyung-Do Kim 3.1.2006
% modified from Brian Harms 6.4.2002
```

```
% Calculates time interval-dependent mean-squared-displacements from
cell
% trajectory coordinates by the method of non-overlapping intervals
(see
% Dickinson and Tranquillo, AIChE J, 1993)

% Code provides the mean-squared displacement data for individual cells
and
% population of cells in the experiment, including the SEM for each
cell's
% MSD to track error propagation through a total experiment cell mean.
It
% also calculates other cell path parameters and plots MSD vs Time
interval

% Store number of time points and number of cells tracked.
nt = size(data,1);
n_cells = size(data,2)./2;

% Counter for total number of displacements per interval length
int_counter = zeros(nt - 1,1);

% Counter for total time points of a track
tp_counter = zeros(1,nt);

% loop to separate the raw data into individual cells
for i = 1:n_cells

    % reads the X and Y coordinates of a single cell
    cell{i} = data(:,2*i-1:2*i);

    % Find all non-zero entries and truncates all zeros off the input
data
    I = find(cell{i}(:,1));
    cell{i} = cell{i}(I,1:2);

    % Add the number of time points of this track to the time-point-
counter
    tp_counter(size(cell{i},1)) = tp_counter(size(cell{i},1)) + 1;

    % x- and y- coordinates cells
    x{i} = cell{i}(:,1);
    y{i} = cell{i}(:,2);
    % Calculate displacements in x and y direction in r.t. first time
point.
    xx{i}= x{i}-x{i}(1);
    yy{i}= y{i}-y{i}(1);

    % Number of possible intervals in current track.
    total_int(i) = length(x{i}) - 1;

    % Vector of number of nonoverlapping intervals for each interval
length (t).
    v_nonoverlapint = floor(total_int(i)./(1:total_int(i))');
```

```
    % Loop for all possible time points for displacement measurements
    for j = 1:total_int(i)

        % Loop all possible starting points for a given time period
        % columns are time values; rows are distances traveled in that
time
        for k = 1:v_nonoverlapint(j)
            distance{i}(k,j) = (x{i}(1 + j*k) - x{i}(1 + j*(k-1)))^2 +
...
                (y{i}(1+j*k) - y{i}(1 + j*(k-1)))^2;
        end

        % Mean squared displacement data for cell{i}
        msd{i}(j,1) = sum(distance{i}(:,j))/v_nonoverlapint(j);
        % Number of intervals for time period i for cell{b}
        msd{i}(j,2) = v_nonoverlapint(j);
        % Same in vector format for later calculations
        msd_v(j,2*i-1) = sum(distance{i}(:,j))/v_nonoverlapint(j);
        msd_v(j,2*i) = v_nonoverlapint(j);

        % Finds non-zero distance measurements for a given cell and
time
        % interval length
        J = find(distance{i}(:,j));

        % Puts these distances in a vector for determination of
variance
        cov_distance{j} = distance{i}(J,j);

        % individual cell SEM for the MSD
        sigma_msd_cell{i}(j) =
cov(cov_distance{j})/sqrt(length(cov_distance{j}));

        % count the total number of intervals for time period j summed
for
        % all cells
        int_counter(j) = int_counter(j) + v_nonoverlapint(j);

    end

    % Calculate path information for this cell
    % Total path length
    total_cell_path_length(i) = (sum((distance{i}(:,1)).^0.5))';
    % Total cell displacement
    total_cell_displacement(i) = (distance{i}(1,total_int(i)))^0.5;
    % Chemotactic Index
    CI(i,1) = total_cell_displacement(i)./total_cell_path_length(i);

end

% Plot MSD vs time interval in a subplot
if n_cells > 25
    columns = 10;
else
```

```
    columns = 5;
end

scrsz = get(0,'ScreenSize');
figure('Position',[scrsz(3)/6 scrsz(4)/6 scrsz(3)/1.4 scrsz(4)/1.4],...
    'Name','MSD vs. Time Plot Window','NumberTitle','off')
rows = ceil(n_cells./columns);
if n_cells < columns
    for i = 1:n_cells
        subplot(1,n_cells,i);
        plot(1:total_int(i),msd{i}(:,1),'xk');
        title(['Cell #',num2str(i)]);
    end
else
    for i = 1:rows
        for j = 1:columns
            if (i - 1).*columns + j > n_cells
                break
            end
            subplot(rows,columns,(i - 1).*columns + j);
            plot(1:total_int((i - 1).*columns + j),msd{(i - 1).*columns
+ j}(:,1),'xk');
            title(['Cell #',num2str((i - 1).*columns + j)]);
        end
    end
end

saveas(gcf,[pwd '\Figures\' expname '_MSDplot.fig']);
saveas(gcf,[pwd '\Tiffs\' expname '_MSDplot'],'tif');

% Calculate mean-square displacements averaged across all cells (weigh
% individual cell by time periods contributed to total)
for i = 1:n_cells
    avg_total_msd(:,i) = msd_v(:,2*i-1).*msd_v(:,2*i)./int_counter;
end
% Average total MSD for each time interval
avg_total_msd = sum(avg_total_msd,2);
% Maximum averaged total MSD
max_msd = max(avg_total_msd);
% prwfit.m - Hyung-Do Kim 3.2.2006
% modified from Lisa Joslin 4.23.2004, originally adapted from Gargi
% Maheshwari

% Fits the mean squared displacement data obtained from msdnonoverlap.m
to
% the persistent random walk model to obtain S and P.  Two methods are
% used: calculate S from MSD then fit P; fit both S and P.  The first
% method usually fares better.
% Plots trajectory, speed at each time point, theoretical curves given
S
% and P, and the MSD scatter
% For pooled data set, the program plots the Wind-Rose plot,
Persistence
% vs. CI plot, S/P plot of each cell, and the histogram of S and P
% distribution.
```

```
global Scalc z
Scalc = [];
Sfit = [];
Pcalc = [];
Pfit = [];

% Fitted P over maximum P counter
Pmisfit = zeros(1,2);

for z = 1:n_cells
    % Convert intervals to hours
    time = interval.*[1:total_int(z)]';

    % Method 1 - Calculate speed and fit persistence

    % Calculate speed from distance values
    Scalc(z,1) = sqrt(msd{z}(1,1))/interval; % Mean

    % Find non-zero values for distances for covariance calculations
    J = find(distance{z}(:,1));
    cov_speed = sqrt(distance{z}(J,1))./interval;

    Scalc(z,2) = std(cov_speed)/sqrt(length(cov_speed));

    % One-parameter least-squares fit for P
    % guess P to be the time interval
    P_0 = interval;
    % Upper and lower bound for P: total duration of exp, 0,
respectively
    [Pcalc(z,1),Calcresnorm(z,1)] = lsqcurvefit(@prwfuncalc, P_0, time,
msd{z}(:,1), 0, nt.*interval,optimset('Display','off'));

    % If P is larger than experiment duration, fitting may be faulty.
    if Pcalc(z) > nt.*interval
        Pcalc(z) = nt.*interval
        Pmisfit(1) = Pmisfit(1) + 1;
    end

    % Method 2 - Fit both speed and persistence
    Par_0 = [5 interval];
    [Par,Fitresnorm(z,1)] = lsqcurvefit(@prwfunfit, Par_0, time,
msd{z}(:,1), [0 0], [1000 nt.*interval],optimset('Display','off'));
    Sfit(z,1) = Par(1);
    Pfit(z,1) = Par(2);

    % If P is larger than experiment duration, fitting may be faulty.
    if Pfit(z) > nt.*interval
        Pfit(z) = nt.*interval
        Pmisfit(2) = Pmisfit(2) + 1;
    end
end

% Plot the fitting of data for checking and trajectory paths and speed
at
% each time point
```

```
% One does not have to see all individual plots:
seeplot = input(['Would you like to look at ' num2str(n_cells) '
individual cell plots? '],'s');
if (seeplot == 'y') || (seeplot == 'Y')
    plotthem = 1;
else
    plotthem = 0;
end

temp_counter = 0;
for i = 1:n_cells
    time = interval.*[1:total_int(i)]';
    % Calculate fitted curves
    model_calc = 2.*Scalc(i,1).^2.*Pcalc(i).*(time - Pcalc(i).*(1 -
exp(-time./Pcalc(i))));
    model_fit = 2.*Sfit(i,1).^2.*Pfit(i).*(time - Pfit(i).*(1 - exp(-
time./Pfit(i))));

    % Plot calculated Speed/fitted Persistence Curve
    figure('Position',[scrsz(3)/6 scrsz(4)/6 scrsz(3)/1.4
scrsz(4)/1.4],...
        'Name',['Model fitting to MSD data - Cell #'
num2str(i)],'NumberTitle','off')
    subplot(2,2,3);
    plot(time,msd{i}(:,1),'xk',time,model_calc,'-r');
    title(['Calc. S = ' num2str(Scalc(i,1)) ' um/h - Fit P = '
num2str(Pcalc(i,1).*60) ' min']);
    xlabel('Time interval [h]');
    ylabel('Mean Squared Displacement [um]');

    % Plot fitted speed/fitted persistence curve
    subplot(2,2,4);
    plot(time,msd{i}(:,1),'xk',time,model_fit,'-r');
    title(['Fit S = ' num2str(Sfit(i,1)) ' um/h - Fit P = '
num2str(Pfit(i,1).*60) ' min']);
    xlabel('Time interval (hr)');
    ylabel('Mean Squared Displacement (um)');

    % Plot cell trajectory
    subplot(2,2,1);
    plot(xx{i}, yy{i},'k.-');
    title('Cell trajectory');
    xlabel('x [um]');
    ylabel('y [um]');

    % Plot speed at each time interval curve
    subplot(2,2,2);
    Splot = sqrt(distance{i}(:,1))./interval;
    plot(time,Splot,'xb');
    title('Speed at each time interval');
    xlabel('Time [h]');
    ylabel('Speed [um/h]');

    saveas(gcf,[pwd '\Figures\' expname '_celln_' num2str(i) '.fig']);
```

```
    saveas(gcf,[pwd '\Tiffs\' expname '_celln_' num2str(i)],'tif');

    if plotthem == 0
        close(gcf)
    else
    % Wait after 10 plots.
        if round(i./10) == i./10
            fprintf(['Cells # ' num2str(temp_counter.*10 + 1) ' - '
num2str(i)]);
            temp_counter = temp_counter + 1;
            wait = input('\nPress Enter to continue','s');
        end
    end
end

% Wait for plots to be looked at
if plotthem == 1
    if round(i./10) == i./10
    else
        fprintf(['Cells # ' num2str(temp_counter.*10 + 1) ' - '
num2str(i)]);
        wait = input('\nPress Enter to continue','s');
    end
end

% Calculate basic statistics of speed and persistence across all cells
speed_mean = mean(Scalc(:,1));
speed_median = median(Scalc(:,1));
speed_sem = std(Scalc(:,1))./sqrt(length(Scalc(:,1)));
speed_max = max(Scalc(:,1));
speed_min = min(Scalc(:,1));
persistence_mean = mean(Pcalc(:,1));
persistence_median = median(Pcalc(:,1));
persistence_sem = std(Pcalc(:,1))./sqrt(length(Pcalc(:,1)));
persistence_max = max(Pcalc(:,1));
persistence_min = min(Pcalc(:,1));
CI_mean = mean(CI(:,1));
CI_median = median(CI(:,1));
CI_sem = std(CI(:,1))./sqrt(length(CI(:,1)));
CI_max = max(CI(:,1));
CI_min = min(CI(:,1));

% Plot Wind-Rose plot of all trajectories
figure;
for i = 1:n_cells
    plot(xx{i},yy{i},'-k');
    hold on;
end
hold off;
title(['Wind-Rose plot: S = ' num2str(speed_mean) ' +- '
num2str(speed_sem)...
    ' um/hr - P = ' num2str(persistence_mean.*60) ' +- '
num2str(persistence_sem.*60) ' min']);
axis([-500 500 -500 500]);
xlabel('x [um]');
ylabel('y [um]');
```

```
saveas(gcf,[pwd '\Figures\' expname '_windrose.fig']);
saveas(gcf,[pwd '\Tiffs\' expname '_windrose'],'tif');

% Visualize Speed and Persistence distribution of all cells
% First, produce lines that show mean speed and persistence
S_Line = zeros(2,1) + speed_mean;
P_Line = zeros(2,1) + persistence_mean.*60;

figure;
% x axis: Speed, y axis: Persistence
plot(Scalc(:,1), Pcalc(:,1).*60, 'ok', S_Line,
linspace(0,persistence_max.*60,2),...
    '-r', linspace(0,speed_max,2), P_Line, '-b');
legend('Single Cells', 'Mean Persistence', 'Mean Speed',2);
xlabel('Speed [um/hr]');
ylabel('Persistence [min]');
axis([0 speed_max 0 persistence_max.*60]);
title('Scatter Plot: Speed and Persistence per cell');

saveas(gcf,[pwd '\Figures\' expname '_SPscatter.fig']);
saveas(gcf,[pwd '\Tiffs\' expname '_SPscatter'],'tif');

% A good way to check persistence values is to compare Chemotactic
Index
% and Persistence Times
figure;
plot(CI,Pcalc(:,1).*60, 'ko');
xlabel('Chemotactic Index');
ylabel('Persistence [min]');
title('Persistence vs. Chemotactic Index Plot');

saveas(gcf,[pwd '\Figures\' expname '_PvsCI.fig']);
saveas(gcf,[pwd '\Tiffs\' expname '_PvsCI'],'tif');

% Generate histograms of speed and persistence
% Parameters for maximum bin edge and bin numbers to make histogram
pretty
if speed_max > 100
    Smax_h = ceil(speed_max./10).*10;
    Sbins = 21;
elseif speed_max > 50
    Smax_h = round(speed_max./10).*10 + 5;
    Sbins = Smax_h./5 + 1;
elseif speed_max > 20
    Smax_h = ceil(ceil(speed_max)./2).*2;
    Sbins = Smax_h./2 + 1;
else
    Smax_h = ceil(speed_max);
    Sbins = Smax_h + 1;
end
if persistence_max.*60 > 100
    Pmax_h = ceil(persistence_max.*60./10).*10;
    Pbins = 20 + 1;
elseif persistence_max.*60 > 50
```

```
    Pmax_h = round(persistence_max.*60./10).*10 + 5;
    Pbins = Pmax_h./5 + 1;
elseif persistence_max.*60 > 20
    Pmax_h = ceil(ceil(persistence_max.*60./2)).*2
    Pbins = Pmax_h./2 + 1;
else
    Pmax_h = ceil(persistence_max.*60);
    Pbins = Pmax_h + 1;
end
if CI_max > 0.5
    CImax_h = ceil(CI_max.*10)./10;
    CIbins = CImax_h./0.05 + 1;
else
    CImax_h = ceil(ceil(CI_max.*100)./2).*2./100;
    CIbins = CImax_h./0.02 + 1;
end
% Create bin vectors
Sbinv = linspace(0,Smax_h,Sbins);
Sgenbinv = linspace(0,150,31);
Pbinv = linspace(0,Pmax_h,Pbins);
Pgenbinv = linspace(0,360,19);
CIbinv = linspace(0,CImax_h,CIbins);
CIgenbinv = linspace(0,1,21);


% Count according to bins
Hist_S = histc(Scalc(:,1),Sbinv)./n_cells;
Histgen_S = histc(Scalc(:,1),Sgenbinv)./n_cells;
Hist_P = histc(Pcalc(:,1).*60,Pbinv)./n_cells;
Histgen_P = histc(Pcalc(:,1).*60,Pgenbinv)./n_cells;
Hist_CI = histc(CI(:,1),CIbinv)./n_cells;
Histgen_CI = histc(CI(:,1),CIgenbinv)./n_cells;

% Plot histogram
figure('Position',[scrsz(3)/6 scrsz(4)/6 scrsz(3)/1.2 scrsz(4)/1.8],...
    'Name','Histogram Plot','NumberTitle','off')
subplot(2,3,1);
bar(Sbinv,Hist_S,'histc');
title('Speed Distribution');
xlabel('Speed [um/h]');
ylabel('Fraction of Total Cells');
axis([0 Smax_h 0 1]);
subplot(2,3,4);
bar(Sgenbinv,Histgen_S,'histc');
xlabel('Speed [um/h]');
ylabel('Fraction of Total Cells');
axis([0 155 0 1]);
subplot(2,3,2);
bar(Pbinv,Hist_P,'histc');
title('Persistence Distribution');
xlabel('Persistence [min]');
ylabel('Fraction of Total Cells');
axis([0 Pmax_h 0 1]);
subplot(2,3,5);
bar(Pgenbinv,Histgen_P,'histc');
xlabel('Persistence [min]');
```

```
ylabel('Fraction of Total Cells');
axis([0 380 0 1]);
subplot(2,3,3);
bar(CIbinv,Hist_CI,'histc');
title('Chemotactic Index Distribution');
xlabel('Chemotactic Index');
ylabel('Fraction of Total Cells');
axis([0 CImax_h 0 1]);
subplot(2,3,6);
bar(CIgenbinv,Histgen_CI,'histc');
xlabel('Chemotactic Index');
ylabel('Fraction of Total Cells');
axis([0 1.05 0 1]);

saveas(gcf,[pwd '\Figures\' expname '_SPCIhist.fig']);
saveas(gcf,[pwd '\Tiffs\' expname '_SPCIhist'],'tif');

% Calculate few more parameters
mu(:,1) = 0.5.*Scalc(:,1).^2.*Pcalc(:,1);
meanpathlength = Scalc(:,1).*Pcalc(:,1);
mu_mean = 0.5.*speed_mean.^2.*persistence_mean;
meanpathlength_mean = speed_mean.*persistence_mean;

%function prwfun %%

function f = prwfun(time, S)

f = 2*S^2*P*[time - P*(1-exp(-time/P))];


%% function prwfuncalc %%
function f = prwfuncalc(P, time)

global Scalc z

f = 2*(Scalc(z,1))^2*P*[time - P*(1-exp(-time/P))];

%% function prwunfit %%

function f = prwfunfit(Par, time)

f = 2*Par(1)^2*Par(2)*[time - Par(2)*(1-exp(-time/Par(2)))];
```

Appendix 2

```
%% Written by Hyung-Do Kim, Lauffenburger Lab %%

function varargout = woundgui(varargin)
% WOUNDGUI M-file for woundgui.fig
%      WOUNDGUI, by itself, creates a new WOUNDGUI or raises the
existing
%      singleton*.
%
%      H = WOUNDGUI returns the handle to a new WOUNDGUI or the handle
to
%      the existing singleton*.
%
%      WOUNDGUI('CALLBACK',hObject,eventData,handles,...) calls the
local
%      function named CALLBACK in WOUNDGUI.M with the given input
arguments.
%
%      WOUNDGUI('Property','Value',...) creates a new WOUNDGUI or
raises the
%      existing singleton*.  Starting from the left, property value
pairs are
%      applied to the GUI before woundgui_OpeningFunction gets called.
An
%      unrecognized property name or invalid value makes property
application
%      stop.  All inputs are passed to woundgui_OpeningFcn via
varargin.
%
%      *See GUI Options on GUIDE's Tools menu.  Choose "GUI allows only
one
%      instance to run (singleton)".
%
% See also: GUIDE, GUIDATA, GUIHANDLES

% Edit the above text to modify the response to help woundgui

% Last Modified by GUIDE v2.5 21-Sep-2005 18:11:03

% Begin initialization code - DO NOT EDIT
gui_Singleton = 1;
gui_State = struct('gui_Name',       mfilename, ...
                   'gui_Singleton',  gui_Singleton, ...
                   'gui_OpeningFcn', @woundgui_OpeningFcn, ...
                   'gui_OutputFcn',  @woundgui_OutputFcn, ...
                   'gui_LayoutFcn',  [] , ...
                   'gui_Callback',   []);
if nargin && ischar(varargin{1})
    gui_State.gui_Callback = str2func(varargin{1});
end

if nargout
    [varargout{1:nargout}] = gui_mainfcn(gui_State, varargin{:});
else
```

```
    gui_mainfcn(gui_State, varargin{:});
end
% End initialization code - DO NOT EDIT


% --- Executes just before woundgui is made visible.
function woundgui_OpeningFcn(hObject, eventdata, handles, varargin)
% This function has no output args, see OutputFcn.
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
% varargin   command line arguments to woundgui (see VARARGIN)

% Choose default command line output for woundgui
handles.output = hObject;

% Update handles structure
guidata(hObject, handles);

% UIWAIT makes woundgui wait for user response (see UIRESUME)
% uiwait(handles.figure1);

handles.load_final = 0;
handles.filetype = '.bmp';
set(handles.workdir,'String',pwd);
set(handles.savedir,'String',pwd);

axes(handles.fig_beg); % Select the proper axes
imshow(ones(512,512));
axes(handles.fig_mid);
imshow(ones(512,512));
axes(handles.fig_end);
imshow(ones(512,512));

addpath(pwd,'-begin');

guidata(hObject, handles);

% --- Outputs from this function are returned to the command line.
function varargout = woundgui_OutputFcn(hObject, eventdata, handles)
% varargout  cell array for returning output args (see VARARGOUT);
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Get default command line output from handles structure
varargout{1} = handles.output;


function workdir_Callback(hObject, eventdata, handles)
% hObject    handle to workdir (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
```

```
% Hints: get(hObject,'String') returns contents of workdir as text
%        str2double(get(hObject,'String')) returns contents of workdir
as a double


% --- Executes during object creation, after setting all properties.
function workdir_CreateFcn(hObject, eventdata, handles)
% hObject    handle to workdir (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end


% --- Executes on button press in workdirbrowse.
function workdirbrowse_Callback(hObject, eventdata, handles)
% hObject    handle to workdirbrowse (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
workdir_text = uigetdir;
if isstr(workdir_text) == 1
    set(handles.workdir,'String',workdir_text);
end


function frames_Callback(hObject, eventdata, handles)
% hObject    handle to frames (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of frames as text
%        str2double(get(hObject,'String')) returns contents of frames
as a double


% --- Executes during object creation, after setting all properties.
function frames_CreateFcn(hObject, eventdata, handles)
% hObject    handle to frames (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end
```

```
function interval_Callback(hObject, eventdata, handles)
% hObject    handle to interval (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of interval as text
%        str2double(get(hObject,'String')) returns contents of interval
as a double


% --- Executes during object creation, after setting all properties.
function interval_CreateFcn(hObject, eventdata, handles)
% hObject    handle to interval (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end


% --- Executes on selection change in imageenc_popup.
function imageenc_popup_Callback(hObject, eventdata, handles)
% hObject    handle to imageenc_popup (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: contents = get(hObject,'String') returns imageenc_popup
contents as cell array
%        contents{get(hObject,'Value')} returns selected item from
imageenc_popup
val = get(hObject,'Value');
switch val
    case 1
        handles.filetype = '.bmp';
    case 2
        handles.filetype = '.tif';
    case 3
        handles.filetype = '.jpg';
end
guidata(hObject,handles);


% --- Executes during object creation, after setting all properties.
function imageenc_popup_CreateFcn(hObject, eventdata, handles)
% hObject    handle to imageenc_popup (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called
```

```
% Hint: popupmenu controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end


% --- Executes on button press in loadbutton.
function loadbutton_Callback(hObject, eventdata, handles)
% hObject    handle to loadbutton (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

cd(get(handles.workdir,'String'));
filename = dir(num2str(['*' handles.filetype]));
handles.nframes = length(filename);

if handles.nframes > 0
    for z = 1:handles.nframes
        [handles.I_orig(:,:,:,z), handles.map] =
imread(filename(z).name);
        handles.I_fill(:,:,:,z) = imfill(handles.I_orig(:,:,:,z));
    end

    set(handles.frames,'String',num2str(handles.nframes));
    pixels = size(handles.I_orig);
    set(handles.pixel_row,'String',num2str(pixels(1)));
    set(handles.pixel_column,'String',num2str(pixels(2)));

    handles.load_final = 1;
    set(handles.message,'String','');

    axes(handles.fig_beg); % Select the proper axes
    imshow(handles.I_fill(:,:,:,1));
    set(handles.frame_beg_n,'String','1');
    axes(handles.fig_mid);
    imshow(handles.I_fill(:,:,:,ceil(handles.nframes./2)));

set(handles.frame_mid_n,'String',num2str(ceil(handles.nframes./2)));
    axes(handles.fig_end);
    imshow(handles.I_fill(:,:,:,handles.nframes));
    set(handles.frame_end_n,'String',num2str(handles.nframes));

    set(handles.frame_beg_n,'String',1);

set(handles.frame_mid_n,'String',num2str(ceil(handles.nframes./2)));
    set(handles.frame_end_n,'String',num2str(handles.nframes));

    set(handles.message,'String','Images Loaded');
    set(handles.savedir,'String',get(handles.workdir,'String'));

    handles.full_update = 0;
    handles.approx_update = 0;
```

```
    handles.linear_final = 0;
    handles.crop_final = 0;
    handles.binary_final = 0;
    handles.linear_restore = 0;
    handles.binary_restore = 0;
else
    set(handles.message,'String',['No ' handles.filetype ' Image Files
Detected']);
end

guidata(hObject,handles);


% --- Executes on button press in restore.
function restore_Callback(hObject, eventdata, handles)
% hObject    handle to restore (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)


% --- Executes on button press in pushbutton7.
function pushbutton7_Callback(hObject, eventdata, handles)
% hObject    handle to pushbutton7 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)


% --- Executes on slider movement.
function thresh_slider_Callback(hObject, eventdata, handles)
% hObject    handle to thresh_slider (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'Value') returns position of slider
%        get(hObject,'Min') and get(hObject,'Max') to determine range
of
%        slider

set(handles.thresh,'String',num2str(get(handles.thresh_slider,'Value'))
);


% --- Executes during object creation, after setting all properties.
function thresh_slider_CreateFcn(hObject, eventdata, handles)
% hObject    handle to thresh_slider (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called

% Hint: slider controls usually have a light gray background.
if isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor',[.9 .9 .9]);
end
```

```
function thresh_Callback(hObject, eventdata, handles)
% hObject    handle to thresh (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of thresh as text
%        str2double(get(hObject,'String')) returns contents of thresh
as a double

val = str2double(get(handles.thresh,'String'));
min = get(handles.thresh,'Min');
max = get(handles.thresh,'Max');
default = 0.15;

if isnumeric(val) & length(val)==1 & val >= min & val <= max
    set(handles.thresh_slider,'Value',val);
elseif isnumeric(val) & length(val)==1 & val <= min
    set(handles.thresh_slider,'Value',min);
    set(handles.thresh,'String',min);
elseif isnumeric(val) & length(val)==1 & val >= max
    set(handles.thresh_slider,'Value',max);
    set(handles.thresh,'String',max);
else
    set(handles.thresh_slider,'Value',default);
    set(handles.thresh,'String',default);
end


% --- Executes during object creation, after setting all properties.
function thresh_CreateFcn(hObject, eventdata, handles)
% hObject    handle to thresh (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end


% --- Executes on button press in begin_crop.
function begin_crop_Callback(hObject, eventdata, handles)
% hObject    handle to begin_crop (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)


% --- Executes on button press in final_crop.
function final_crop_Callback(hObject, eventdata, handles)
```

```
% hObject    handle to final_crop (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)


% --- Executes on button press in skip_crop.
function skip_crop_Callback(hObject, eventdata, handles)
% hObject    handle to skip_crop (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

if handles.binary_final == 1;
    handles.I_crop = handles.I_final;
    handles.crop_final = 1;
else
    set(handles.message,'String','Binarization Missing');
end

guidata(hObject,handles);


function savedir_Callback(hObject, eventdata, handles)
% hObject    handle to savedir (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of savedir as text
%        str2double(get(hObject,'String')) returns contents of savedir
as a double


% --- Executes during object creation, after setting all properties.
function savedir_CreateFcn(hObject, eventdata, handles)
% hObject    handle to savedir (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end


% --- Executes on button press in savedir_browse.
function savedir_browse_Callback(hObject, eventdata, handles)
% hObject    handle to savedir_browse (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
savedir_text = uigetdir;
if isstr(savedir_text) == 1
    set(handles.savedir,'String',savedir_text);
```

```
end

function pixel_row_Callback(hObject, eventdata, handles)
% hObject    handle to pixel_row (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of pixel_row as text
%        str2double(get(hObject,'String')) returns contents of
pixel_row as a double


% --- Executes during object creation, after setting all properties.
function pixel_row_CreateFcn(hObject, eventdata, handles)
% hObject    handle to pixel_row (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end



function pixel_column_Callback(hObject, eventdata, handles)
% hObject    handle to pixel_column (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of pixel_column as text
%        str2double(get(hObject,'String')) returns contents of
pixel_column as a double


% --- Executes during object creation, after setting all properties.
function pixel_column_CreateFcn(hObject, eventdata, handles)
% hObject    handle to pixel_column (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end


function csvname_Callback(hObject, eventdata, handles)
```

```
% hObject    handle to csvname (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of csvname as text
%        str2double(get(hObject,'String')) returns contents of csvname
as a double


% --- Executes during object creation, after setting all properties.
function csvname_CreateFcn(hObject, eventdata, handles)
% hObject    handle to csvname (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end


% --- Executes on button press in savebutton.
function savebutton_Callback(hObject, eventdata, handles)
% hObject    handle to savebutton (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

set(handles.message,'String','');
time_format = 0;

if handles.crop_final == 1
    pixels = size(handles.I_crop);
    totalarea = pixels(1).*pixels(2);

    for z = 1:handles.nframes
        handles.area(z) = totalarea - bwarea(handles.I_crop(:,:,:,z));
    end

    if isempty(get(handles.csvname,'String')) == 1
        set(handles.csvname,'String','data');
    end

    figure(1);
    orig_movie = immovie(handles.I_orig,handles.map);
    figure(1);
    bw_movie = immovie(handles.I_final,handles.map);
    movie2avi(orig_movie,[get(handles.savedir,'String') '\'
get(handles.csvname,'String') '_orig' '.avi'],'compression','none');
    movie2avi(bw_movie,[get(handles.savedir,'String') '\'
get(handles.csvname,'String') '_bw' '.avi'],'compression','none');
```

```
    x = linspace(0,handles.nframes - 1,handles.nframes);

    val = str2double(get(handles.interval,'String'));
    if isnumeric(val) & length(val)==1
        if val > 0
            time_format = 1;
            time = x.*val;
        end
    else
        set(handles.interval,'String','');
    end

    figure(1);
    if time_format == 1
        plot(time,handles.area./handles.area(1));
        title('Normalized Wound Area vs. Time');
        xlabel('Time [min]');
        ylabel('Normalized Wound Area');
    else
        plot(x,handles.area./handles.area(1));
        title('Normalized Wound Area vs. Frame');
        xlabel('Frame #');
        ylabel('Normalized Wound Area');
    end

    % Export data in a csv-file.

    if time_format == 1
        export(:,1) = x';
        export(:,2) = time';
        export(:,3) = handles.area'./handles.area(1);
    else
        export(:,1) = x';
        export(:,2) = handles.area'./handles.area(1);
    end

    expfile = [get(handles.savedir,'String') '\'
get(handles.csvname,'String') '.csv'];
    dlmwrite(expfile,export,',');
    set(handles.message,'String','Data Successfully Saved');
else
    set(handles.message,'String','Image Cropping Missing');
end



% --- Executes on button press in approx.
function approx_Callback(hObject, eventdata, handles)
% hObject    handle to approx (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hint: get(hObject,'Value') returns toggle state of approx
if get(hObject,'Value') == 1;
    set(handles.full,'Value',0);
```

```
end
if get(hObject,'Value') == 0 && get(handles.full,'Value') == 0
    set(handles.approx,'Value',1);
end

% --- Executes on button press in full.
function full_Callback(hObject, eventdata, handles)
% hObject    handle to full (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hint: get(hObject,'Value') returns toggle state of full
if get(hObject,'Value') == 1;
    set(handles.approx,'Value',0);
end
if get(hObject,'Value') == 0 && get(handles.approx,'Value') == 0
    set(handles.full,'Value',1);
end

% --- Executes on button press in update_2.
function update_2_Callback(hObject, eventdata, handles)
% hObject    handle to update_2 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

if handles.full_update == 0 && get(handles.full,'Value') == 1 &&
get(handles.approx,'Value') == 0 && handles.load_final == 1
    % Background subraction full
    for z = 1:handles.nframes
        background(:,:,:,z) =
imopen(handles.I_fill(:,:,:,z),strel('disk',100,4));
        background2(:,:,:,z) = imadjust(background(:,:,:,z),[],[0
0.7]);
        I_back(:,:,:,z) =
imsubtract(handles.I_fill(:,:,:,z),background2(:,:,:,z));
        handles.I_contrast(:,:,:,z) = imadjust(I_back(:,:,:,z));
        n = ceil(100.*z./handles.nframes);
    end
    % set(handles.progress,'String',' ');
    % set(handles.percent,'String',' ');
    handles.full_update = 1;
    set(handles.message,'String','Background Subtraction Complete');
elseif handles.approx_update == 0 && get(handles.full,'Value') == 0 &&
get(handles.approx,'Value') == 1  && handles.load_final == 1
    background =
imopen(handles.I_fill(:,:,:,handles.nframes),strel('disk',100,4));
    background2 = imadjust(background,[],[0 0.7]);
    for z = 1:handles.nframes
        I_back(:,:,:,z) =
imsubtract(handles.I_fill(:,:,:,z),background2);
        handles.I_contrast(:,:,:,z) = imadjust(I_back(:,:,:,z));
        n = ceil(100.*z./handles.nframes);
    end
    % set(handles.progress,'String',' ');
    % set(handles.percent,'String',' ');
    handles.approx_update = 1;
```

```
    set(handles.message,'String','Background Subtraction Complete');
else
    set(handles.message,'String','Load Image Files First');
end

if handles.load_final == 1
    axes(handles.fig_beg); % Select the proper axes

imshow(handles.I_contrast(:,:,:,str2num(get(handles.frame_beg_n,'String
')))));
    axes(handles.fig_mid);

imshow(handles.I_contrast(:,:,:,str2num(get(handles.frame_mid_n,'String
')))));
    axes(handles.fig_end);

imshow(handles.I_contrast(:,:,:,str2num(get(handles.frame_end_n,'String
')))));
end

guidata(hObject,handles);

% --- Executes on button press in update_3.
function update_3_Callback(hObject, eventdata, handles)
% hObject    handle to update_3 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

if handles.full_update == 1 || handles.approx_update == 1
    for z = 1:handles.nframes
        I_filter(:,:,:,z) = imadjust(handles.I_contrast(:,:,:,z),[0
get(handles.index_slider,'Value')],[0
1],get(handles.gamma_slider,'Value'));
        handles.I_noise(:,:,:,z) = medfilt2(I_filter(:,:,:,z),[5 5]);
    end
    set(handles.message,'String','');

    axes(handles.fig_beg); % Select the proper axes

imshow(handles.I_noise(:,:,:,str2num(get(handles.frame_beg_n,'String'))
));
    axes(handles.fig_mid);

imshow(handles.I_noise(:,:,:,str2num(get(handles.frame_mid_n,'String'))
));
    axes(handles.fig_end);

imshow(handles.I_noise(:,:,:,str2num(get(handles.frame_end_n,'String'))
));
    handles.linear_final = 1;
    handles.linear_restore = 0;
    set(handles.message,'String','Linear Scaling Complete');
else
    set(handles.message,'String','Background Subtraction Missing');
end
```

```
guidata(hObject,handles);


% --- Executes on button press in restore_3.
function restore_3_Callback(hObject, eventdata, handles)
% hObject    handle to restore_3 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

if handles.full_update == 1 || handles.approx_update == 1
    axes(handles.fig_beg); % Select the proper axes

imshow(handles.I_contrast(:,:,:,str2num(get(handles.frame_beg_n,'String
'))));
    axes(handles.fig_mid);

imshow(handles.I_contrast(:,:,:,str2num(get(handles.frame_mid_n,'String
'))));
    axes(handles.fig_end);

imshow(handles.I_contrast(:,:,:,str2num(get(handles.frame_end_n,'String
'))));
    set(handles.message,'String','');
    handles.linear_restore = 1;
else
    set(handles.message,'String','Background Subtraction Missing');
end
guidata(hObject,handles);


% --- Executes on slider movement.
function index_slider_Callback(hObject, eventdata, handles)
% hObject    handle to index_slider (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'Value') returns position of slider
%        get(hObject,'Min') and get(hObject,'Max') to determine range
of slider

set(handles.index,'String',num2str(get(handles.index_slider,'Value')));


% --- Executes during object creation, after setting all properties.
function index_slider_CreateFcn(hObject, eventdata, handles)
% hObject    handle to index_slider (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called

% Hint: slider controls usually have a light gray background.
if isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor',[.9 .9 .9]);
end
```

```
function index_Callback(hObject, eventdata, handles)
% hObject    handle to index (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of index as text
%        str2double(get(hObject,'String')) returns contents of index as
a double

val = str2double(get(handles.index,'String'));
min = get(handles.index,'Min');
max = get(handles.index,'Max');
default = 0.5;

if isnumeric(val) & length(val)==1 & val >= min & val <= max
    set(handles.index_slider,'Value',val);
    set(handles.index,'String',num2str(val));
elseif isnumeric(val) & length(val)==1 & val <= min
    set(handles.index_slider,'Value',min);
    set(handles.index,'String','0.0');
elseif isnumeric(val) & length(val)==1 & val >= max
    set(handles.index_slider,'Value',max);
    set(handles.index,'String','1.0');
else
    set(handles.index_slider,'Value',default);
    set(handles.index,'String',default);
end

% --- Executes during object creation, after setting all properties.
function index_CreateFcn(hObject, eventdata, handles)
% hObject    handle to index (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end

% --- Executes on slider movement.
function gamma_slider_Callback(hObject, eventdata, handles)
% hObject    handle to gamma_slider (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'Value') returns position of slider
%        get(hObject,'Min') and get(hObject,'Max') to determine range
of slider

set(handles.gamma,'String',num2str(get(handles.gamma_slider,'Value')));
```

```
% --- Executes during object creation, after setting all properties.
function gamma_slider_CreateFcn(hObject, eventdata, handles)
% hObject    handle to gamma_slider (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called

% Hint: slider controls usually have a light gray background.
if isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor',[.9 .9 .9]);
end


function gamma_Callback(hObject, eventdata, handles)
% hObject    handle to gamma (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of gamma as text
%        str2double(get(hObject,'String')) returns contents of gamma as
a double

val = str2double(get(handles.gamma,'String'));
min = get(handles.gamma,'Min');
max = get(handles.gamma,'Max');
default = 0.7;

if isnumeric(val) & length(val)==1 & val >= min & val <= max
    set(handles.gamma_slider,'Value',val);
    set(handles.gamma,'String',num2str(val));
elseif isnumeric(val) & length(val)==1 & val <= min
    set(handles.gamma_slider,'Value',min);
    set(handles.gamma,'String','0.0');
elseif isnumeric(val) & length(val)==1 & val >= max
    set(handles.gamma_slider,'Value',max);
    set(handles.gamma,'String','1.0');
else
    set(handles.gamma_slider,'Value',default);
    set(handles.gamma,'String',default);
end


% --- Executes during object creation, after setting all properties.
function gamma_CreateFcn(hObject, eventdata, handles)
% hObject    handle to gamma (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
```

```
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end

% --- Executes on button press in update_5.
function update_5_Callback(hObject, eventdata, handles)
% hObject    handle to update_5 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Specify Image Region of Interest
set(handles.message,'String','');

if handles.binary_final == 1
    rect = getrect(handles.fig_beg);
    axes(handles.fig_beg);


rectangle('Position',[rect(1),rect(2),rect(3),rect(4)],'LineWidth',2,'E
dgeColor',[1 0 0]);
    axes(handles.fig_mid);

rectangle('Position',[rect(1),rect(2),rect(3),rect(4)],'LineWidth',2,'E
dgeColor',[1 0 0]);
    axes(handles.fig_end);

rectangle('Position',[rect(1),rect(2),rect(3),rect(4)],'LineWidth',2,'E
dgeColor',[1 0 0]);

    handles.coord = rect;
else
    set(handles.message,'String','Binarization Missing');
end

guidata(hObject,handles);


% --- Executes on button press in restore_5.
function restore_5_Callback(hObject, eventdata, handles)
% hObject    handle to restore_5 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

if handles.binary_final == 1
    axes(handles.fig_beg); % Select the proper axes

imshow(handles.I_final(:,:,:,str2num(get(handles.frame_beg_n,'String'))
));
    axes(handles.fig_mid);

imshow(handles.I_final(:,:,:,str2num(get(handles.frame_mid_n,'String'))
));
    axes(handles.fig_end);
```

```
imshow(handles.I_final(:,:,:,str2num(get(handles.frame_end_n,'String'))
));
else
    set(handles.message,'String','Binarization Missing');
end


% --- Executes on button press in final_5.
function final_5_Callback(hObject, eventdata, handles)
% hObject    handle to final_5 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

handles.I_crop = [];

if handles.binary_final == 1
    for z = 1:handles.nframes
        handles.I_crop(:,:,:,z) =
imcrop(handles.I_final(:,:,:,z),handles.coord);
    end
    axes(handles.fig_beg); % Select the proper axes

imshow(handles.I_crop(:,:,:,str2num(get(handles.frame_beg_n,'String')))
);
    axes(handles.fig_mid);

imshow(handles.I_crop(:,:,:,str2num(get(handles.frame_mid_n,'String')))
);
    axes(handles.fig_end);

imshow(handles.I_crop(:,:,:,str2num(get(handles.frame_end_n,'String')))
);

    handles.crop_final = 1;
else
    set(handles.message,'String','Binarization Missing');
end

guidata(hObject,handles);


% --- Executes on button press in update_4.
function update_4_Callback(hObject, eventdata, handles)
% hObject    handle to update_4 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Binarization
if handles.linear_final == 1;
    for z = 1:handles.nframes
        level = graythresh(handles.I_noise(:,:,:,z)) -
get(handles.thresh_slider,'Value');
        I_bw(:,:,:,z) = im2bw(handles.I_noise(:,:,:,z),level);
        I_erode(:,:,:,z) = bwmorph(I_bw(:,:,:,z),'erode');
```

```
        handles.I_final(:,:,:,z) =
bwmorph(I_erode(:,:,:,z),'dilate',2);
    end
    axes(handles.fig_beg); % Select the proper axes

imshow(handles.I_final(:,:,:,str2num(get(handles.frame_beg_n,'String'))
));
    axes(handles.fig_mid);

imshow(handles.I_final(:,:,:,str2num(get(handles.frame_mid_n,'String'))
));
    axes(handles.fig_end);

imshow(handles.I_final(:,:,:,str2num(get(handles.frame_end_n,'String'))
));
    handles.binary_final = 1;
    handles.binary_restore = 0;
    set(handles.message,'String','Binarization Complete');
else
    set(handles.message,'String','Linear Scaling Missing');
end

guidata(hObject,handles);


% --- Executes on button press in restore_4.
function restore_4_Callback(hObject, eventdata, handles)
% hObject    handle to restore_4 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

if handles.linear_final == 1;
    axes(handles.fig_beg); % Select the proper axes

imshow(handles.I_noise(:,:,:,str2num(get(handles.frame_beg_n,'String'))
));
    axes(handles.fig_mid);

imshow(handles.I_noise(:,:,:,str2num(get(handles.frame_mid_n,'String'))
));
    axes(handles.fig_end);

imshow(handles.I_noise(:,:,:,str2num(get(handles.frame_end_n,'String'))
));
    set(handles.message,'String','');
    handles.binary_restore = 1;
else
    set(handles.message,'String','Linear Scaling Missing');
end

guidata(hObject,handles);


% --- Executes on button press in automate.
function automate_Callback(hObject, eventdata, handles)
```

```
% hObject    handle to automate (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hint: get(hObject,'Value') returns toggle state of automate

% No Background subtraction done yet.

if handles.load_final == 1
    if handles.full_update == 0 && handles.approx_update == 0
        if get(handles.full,'Value') == 1 &&
get(handles.approx,'Value') == 0
            % Background subraction full
            for z = 1:handles.nframes
                background(:,:,:,z) =
imopen(handles.I_fill(:,:,:,z),strel('disk',100,4));
                background2(:,:,:,z) =
imadjust(background(:,:,:,z),[],[0 0.7]);
                I_back(:,:,:,z) =
imsubtract(handles.I_fill(:,:,:,z),background2(:,:,:,z));
                handles.I_contrast(:,:,:,z) =
imadjust(I_back(:,:,:,z));
                n = ceil(100.*z./handles.nframes);
            end
            % set(handles.progress,'String',' ');
            % set(handles.percent,'String',' ');
            handles.full_update = 1;
            set(handles.message,'String','Background Subtraction
Complete');
        elseif handles.approx_update == 0 && get(handles.full,'Value')
== 0 && get(handles.approx,'Value') == 1
            background =
imopen(handles.I_fill(:,:,:,handles.nframes),strel('disk',100,4));
            background2 = imadjust(background,[],[0 0.7]);
            for z = 1:handles.nframes
                I_back(:,:,:,z) =
imsubtract(handles.I_fill(:,:,:,z),background2);
                handles.I_contrast(:,:,:,z) =
imadjust(I_back(:,:,:,z));
                n = ceil(100.*z./handles.nframes);
            end
            % set(handles.progress,'String',' ');
            % set(handles.percent,'String',' ');
            handles.approx_update = 1;
            set(handles.message,'String','Background Subtraction
Complete');
        end
    end
    if handles.linear_final == 0
        for z = 1:handles.nframes
            I_filter(:,:,:,z) = imadjust(handles.I_contrast(:,:,:,z),[0
get(handles.index_slider,'Value')],[0
1],get(handles.gamma_slider,'Value'));
            handles.I_noise(:,:,:,z) = medfilt2(I_filter(:,:,:,z),[5
5]);
        end
```

```
        handles.linear_final = 1;
        set(handles.message,'String','Linear Scaling Complete');
    end
    if handles.binary_final == 0
        for z = 1:handles.nframes
            level = graythresh(handles.I_noise(:,:,:,z)) -
get(handles.thresh_slider,'Value');
            I_bw(:,:,:,z) = im2bw(handles.I_noise(:,:,:,z),level);
            I_erode(:,:,:,z) = bwmorph(I_bw(:,:,:,z),'erode');
            handles.I_final(:,:,:,z) =
bwmorph(I_erode(:,:,:,z),'dilate',2);
        end
        handles.binary_final = 1;
        set(handles.message,'String','Binarization Complete');
    end
    axes(handles.fig_beg); % Select the proper axes

imshow(handles.I_final(:,:,:,str2num(get(handles.frame_beg_n,'String'))
));
    axes(handles.fig_mid);

imshow(handles.I_final(:,:,:,str2num(get(handles.frame_mid_n,'String'))
));
    axes(handles.fig_end);

imshow(handles.I_final(:,:,:,str2num(get(handles.frame_end_n,'String'))
));
else
    set(handles.message,'String','Load Image Files First');
end

guidata(hObject,handles);


function frame_beg_n_Callback(hObject, eventdata, handles)
% hObject    handle to frame_beg_n (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of frame_beg_n as text
%        str2double(get(hObject,'String')) returns contents of
frame_beg_n as a double

if handles.full_update == 0 && handles.approx_update == 0
    axes(handles.fig_beg); % Select the proper axes

imshow(handles.I_fill(:,:,:,str2num(get(handles.frame_beg_n,'String')))
);
elseif handles.linear_final == 0 || handles.linear_restore == 1
    axes(handles.fig_beg); % Select the proper axes

imshow(handles.I_contrast(:,:,:,str2num(get(handles.frame_beg_n,'String
'))));
elseif handles.binary_final == 0 || handles.binary_restore == 1
    axes(handles.fig_beg); % Select the proper axes
```

```
imshow(handles.I_noise(:,:,:,str2num(get(handles.frame_beg_n,'String'))
));
elseif handles.binary_final == 1
    axes(handles.fig_beg);

imshow(handles.I_final(:,:,:,str2num(get(handles.frame_beg_n,'String'))
));
end


% --- Executes during object creation, after setting all properties.
function frame_beg_n_CreateFcn(hObject, eventdata, handles)
% hObject    handle to frame_beg_n (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end


function frame_mid_n_Callback(hObject, eventdata, handles)
% hObject    handle to frame_mid_n (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of frame_mid_n as text
%        str2double(get(hObject,'String')) returns contents of
frame_mid_n as a double

if handles.full_update == 0 && handles.approx_update == 0
    axes(handles.fig_mid); % Select the proper axes

imshow(handles.I_fill(:,:,:,str2num(get(handles.frame_mid_n,'String')))
);
elseif handles.linear_final == 0 || handles.linear_restore == 1
    axes(handles.fig_mid); % Select the proper axes

imshow(handles.I_contrast(:,:,:,str2num(get(handles.frame_mid_n,'String
')))));
elseif handles.binary_final == 0 || handles.binary_restore == 1
    axes(handles.fig_mid); % Select the proper axes

imshow(handles.I_noise(:,:,:,str2num(get(handles.frame_mid_n,'String'))
));
elseif handles.binary_final == 1
    axes(handles.fig_mid);

imshow(handles.I_final(:,:,:,str2num(get(handles.frame_mid_n,'String'))
));
```

```
end


% --- Executes during object creation, after setting all properties.
function frame_mid_n_CreateFcn(hObject, eventdata, handles)
% hObject    handle to frame_mid_n (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end


function frame_end_n_Callback(hObject, eventdata, handles)
% hObject    handle to frame_end_n (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of frame_end_n as text
%        str2double(get(hObject,'String')) returns contents of
frame_end_n as a double

if handles.full_update == 0 && handles.approx_update == 0
    axes(handles.fig_end); % Select the proper axes

imshow(handles.I_fill(:,:,:,str2num(get(handles.frame_end_n,'String')))
);
elseif handles.linear_final == 0 || handles.linear_restore == 1
    axes(handles.fig_end); % Select the proper axes

imshow(handles.I_contrast(:,:,:,str2num(get(handles.frame_end_n,'String
')))));
elseif handles.binary_final == 0 || handles.binary_restore == 1
    axes(handles.fig_end); % Select the proper axes

imshow(handles.I_noise(:,:,:,str2num(get(handles.frame_end_n,'String'))
));
elseif handles.binary_final == 1
    axes(handles.fig_end);

imshow(handles.I_final(:,:,:,str2num(get(handles.frame_end_n,'String'))
));
end

% --- Executes during object creation, after setting all properties.
function frame_end_n_CreateFcn(hObject, eventdata, handles)
% hObject    handle to frame_end_n (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called
```

```
% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end
```

Appendix 3

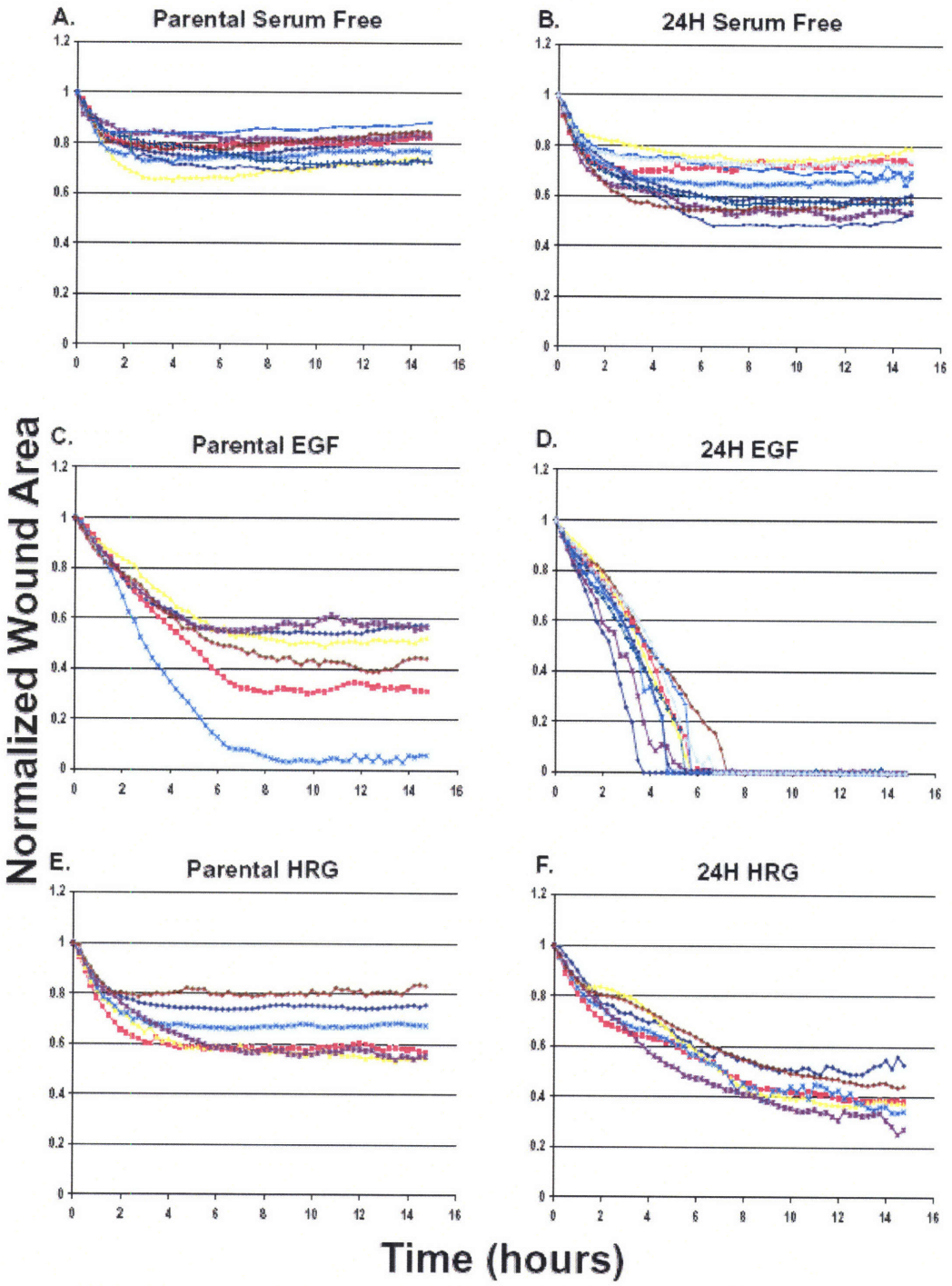


Figure A3-1: Wound closure curves for all wells. Normalized wound area measured every 15 minutes is reported for all wells observed under each condition.

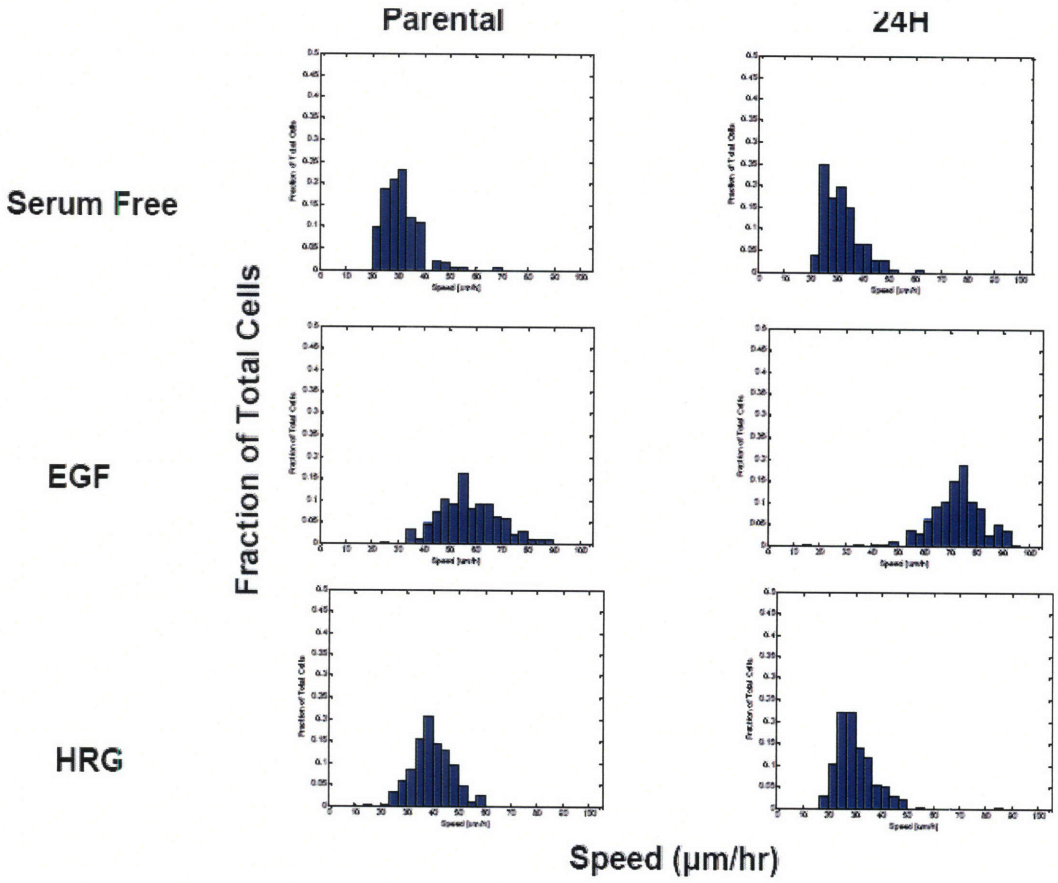


Figure A3-2: Histogram of raw data: cell speed (μm/hr). Cell speed distributions for 24H and parental cells treated with EGF (100 ng.ml), HRG (80 ng/ml), or under serum-free conditions. Total cell number vary from 153-196.

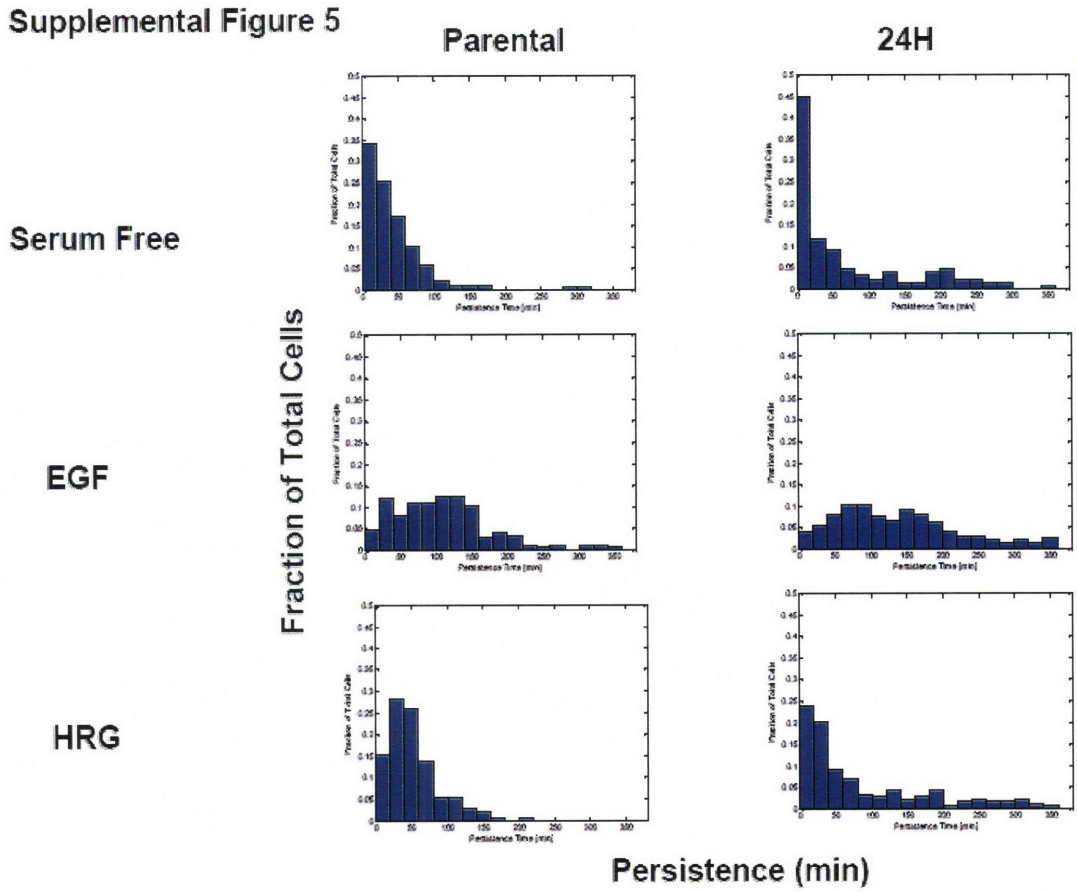


Figure A3-3: Histogram of raw data: cell directional persistence (min). Cell persistence distributions for 24H and parental cells treated with EGF (100 ng.ml), HRG (80 ng/ml), or under serum-free conditions. Total cell number vary from 153-196.

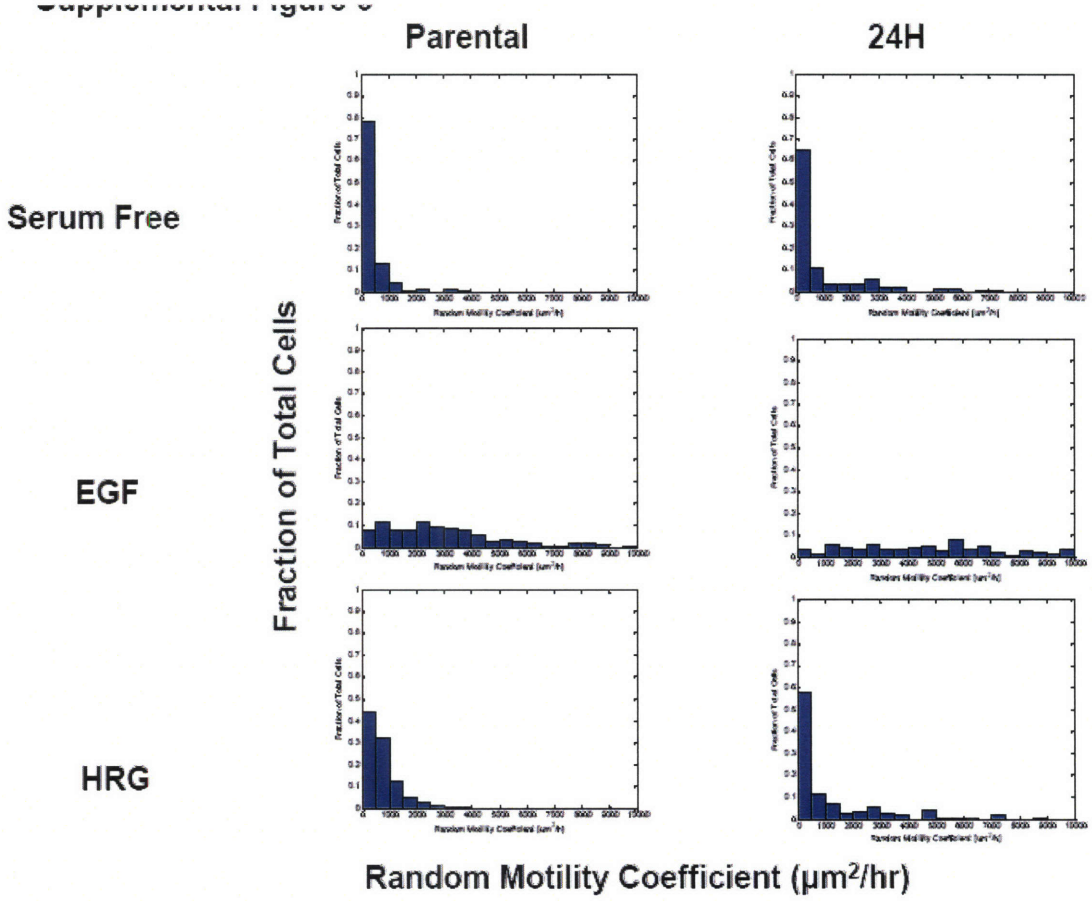


Figure A3-4: Histogram of raw data: cell random motility coefficient (μm2/hr). Cell random motility coefficient distributions for 24H and parental cells treated with EGF (100 ng/ml), HRG (80 ng/ml), or under serum-free conditions. Total cell number vary from 153-196.

Appendix 4

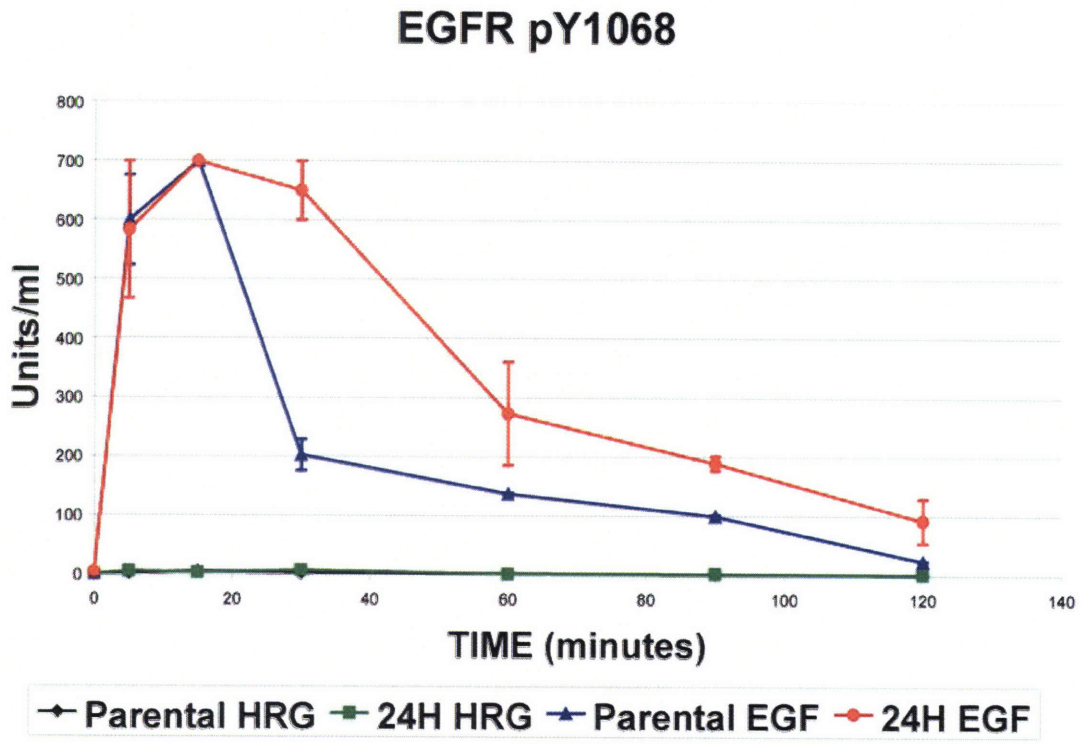


Figure A4-1 : EGFR Y1068 phosphorylation in parental and 24H cells. Cells were simulated with both HRG (80 ng/ml) and EGF (100 ng/ml). All data shown ± SEM.

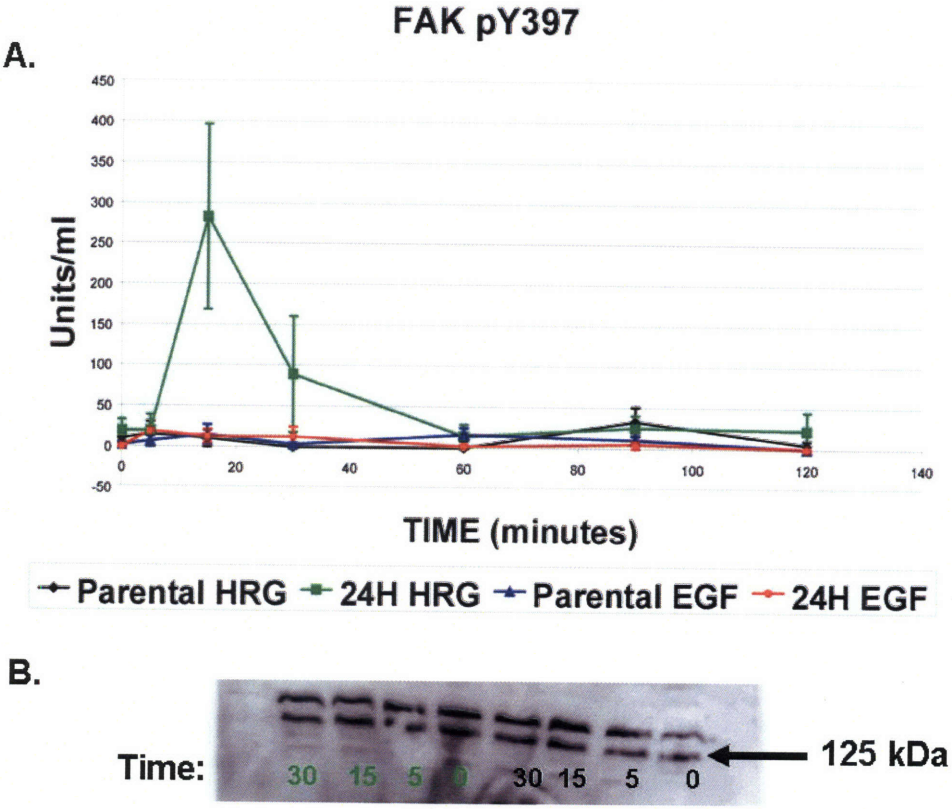


Figure A4-2 : Fak phosphorylation in parental and 24H cells. Cells were simulated with both HRG (80 ng/ml) and EGF (100 ng/ml). All data shown ± SEM.

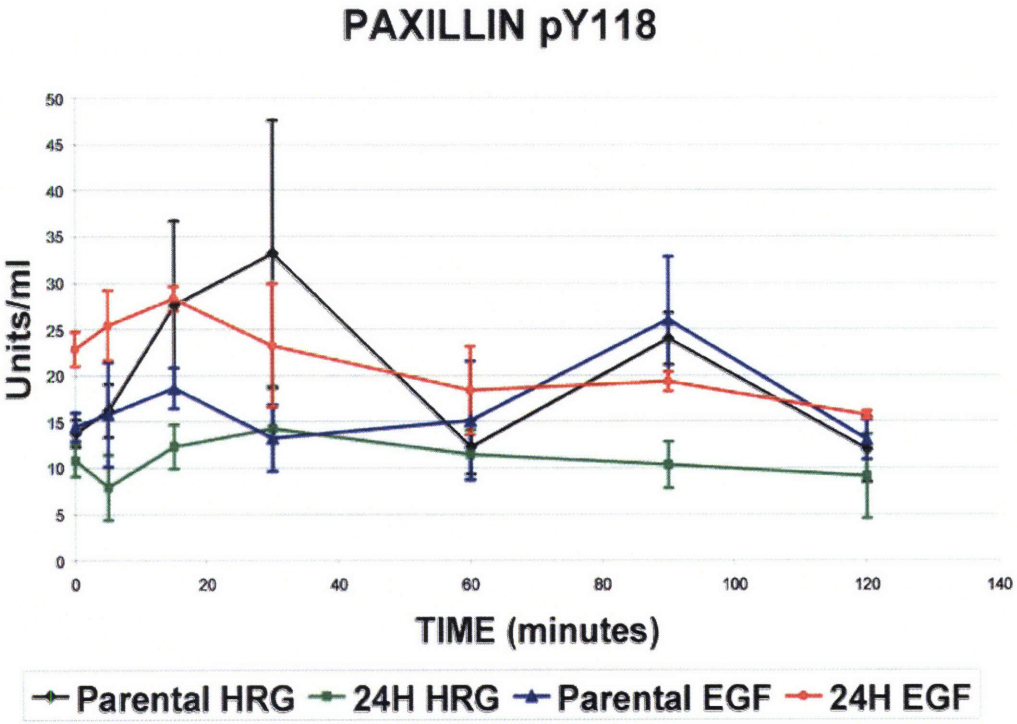


Figure A4-3 : Paxillin phosphorylation in parental and 24H cells. Cells were simulated with both HRG (80 ng/ml) and EGF (100 ng/ml). All data shown ± SEM.

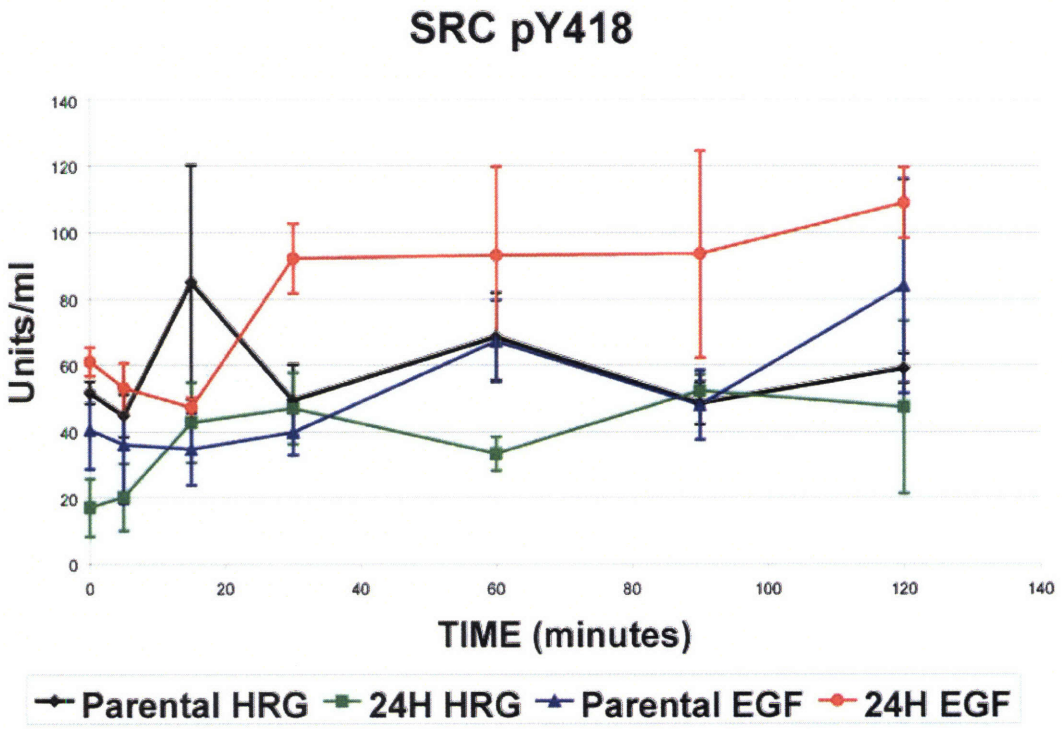


Figure A4-4: Src phosphorylation in parental and 24H cells. Cells were simulated with both HRG (80 ng/ml) and EGF (100 ng/ml). All data shown ± SEM.