Understanding and Targeting Network-Level Sheddase Regulation in Invasive Disease

by

Miles Aaron Miller

A.B., Princeton University (2008)

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

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2013

Author .....
Department of Biological Engineering
May 3, 2013

Certified by .....
Douglas A. Lauffenburger
Professor
Thesis Supervisor

Certified by .....
Linda G. Griffith
Professor
Thesis Supervisor

Accepted by .....
Forest M. White
Chairman, Graduate Program Committee

Thesis Committee

Alan J. Grodzinsky, Professor (Thesis Committee Chair)

Harvey F. Lodish, Professor

Frank B. Gertler, Professor

Understanding and Targeting Network-Level Sheddase Regulation in Invasive Disease

by

Miles Aaron Miller

Submitted to the Department of Biological Engineering
on May 3, 2013, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Regulated cell-surface proteolysis underpins key processes of cellular growth and motility in both physiological and pathological contexts. However, comprehending how multiple proteolytic events cohesively integrate to yield context-dependent cellular behavior remains a challenge in the fields of both protease biology and systems biology in general. This work begins to address that challenge by quantitatively investigating the integrated effect of multiple diverse proteolytic events and their interaction with cell-signaling pathways from a computational network perspective, particularly focusing on A Disintegrin and Metalloproteinases (ADAMs). ADAMs have been studied for decades as the principal cell-surface "sheddases" responsible for cleaving growth factor ligands and receptor tyrosine kinase ectodomains from the cell surface. However, activity regulation, feedback, and catalytic promiscuity impede our understanding of context-dependent sheddase function, and clinical trials targeting metalloproteinases in cancer have failed in part due to a poor understanding of the complex functions they mediate.

This thesis outlines a conceptual framework for studying protease network biology (Chapter 1), describes novel experimental methods designed for such a framework (Chapters 2-3), and applies both to understand protease regulation in invasive disease (Chapter 4). Using combined measurement and computational modeling, we present a paradigm for monitoring and analyzing complex networks of protease activities that interface with signaling pathways to influence cellular migration in the invasive diseases of cancer and endometriosis. We find sheddase activity integrates with signaling pathways to direct cell migration, especially through concomitant proteolysis of both ligands and receptors. We find that indirect reduction of sheddase activity through kinase inhibition can lead to an accumulation of growth-factor receptors on the cell surface, consequently producing undesired compensatory signaling feedback. Thus, here we present a novel mechanism of rapid, protease-driven resistance to kinase inhibitors, and we subsequently demonstrate strategies for overcoming resistance through drug combinations. We develop a novel microfluidic platform to study protease activities in clinical samples, and apply the technology to study the

peritoneal fluid from endometriosis patients. Results indicate joint dysregulation of sheddase activity with disease. Overall, this work provides a model for measuring, understanding, and targeting networks of proteases and the kinases with which they interact.

Thesis Supervisor: Douglas A. Lauffenburger
Title: Professor

Thesis Supervisor: Linda G. Griffith
Title: Professor

Acknowledgments

I have many people to thank for help and support in the completion of my thesis. My graduate advisors, Douglas Lauffenburger and Linda Griffith, have been truly wonderful mentors and I thank them for their thoughtful scientific and personal guidance. I am furthermore grateful for the time and attention they have generously given me and my work, with many insightful comments, recommendations, and edits over the years.

I owe gratitude to several other faculty who have helped me along the way. My thesis committee, consisting of Harvey Lodish, Frank Gertler, and chaired by Alan Grodzinsky, has been especially helpful in guiding the direction of the thesis and providing feedback and suggestions. In addition to closely working with members of my thesis committee, I am lucky to have directly collaborated with several faculty: Jongyoon Han, Ron Weiss, Steven Tannenbaum, Keith Isaacson (Harvard Medical School), Andreas Herrlich (Harvard Medical School), Chia-Hung Chen (National University of Singapore), Joerg Bartsch (University of Marburg), and Nikola Vujanovic (University of Pittsburgh). These interactions and collaborations have been enjoyable, educational, and I believe profitable to the success and completion of this thesis. I also thank Marcia Moss and Fred Rasmussen (Biozyme, Inc.) for generously contributing substantial time, resources, and expertise in working with me on this thesis. Barbara Imperiali, Marsha Moses (Harvard Medical School), Manu Platt (Georgia Institute of Technology), Kevin Osteen (Vanderbilt University), Stephen Palmer (EMD-Serono), H. Steven Wiley (Pacific Northwest National Laboratory), and Carl Blobel (Weill-Cornell) have given their time through helpful discussions and suggestions, as well as materials and reagents, and for that I am thankful.

My lab-mates have been fantastic colleagues and friends over the years, and I owe them immense thanks. In particular, I thank Aaron Meyer and Chia-Hung Chen for close collaborations and extensive contributions to writing and experimentation. I also express gratitude towards Michael Beste, Aniruddh Sarkar, Caroline Chopko, Shannon Hughes, Madeleine Oudin, Allison Claas, and Ravi Kodihalli for being great collaborators. Additionally, I owe a special thanks to my fellow office-mates not yet mentioned, Jorge Valdez Macias, Lu Alvarez, Linda Stockdale, Kara Huang, and Simon Gordonov, for being good friends and making a great work environment. I also thank Joel Wagner for helpful comments on manuscripts and presentations, and Hsinhwa Lee for being such an outstanding lab manager.

Finally, I would like to acknowledge financial assistance from the Viterbi Fellowship Fund, the National Science Foundation, and the National Institutes of Health.

Dedication

I dedicate this thesis to my fiancee, Karen; my parents, Mike and Anne; and my sister, Kelsey, for being limitless sources of support, love, and inspiration.

Contents

List of Figures

List of Tables

Chapter 1

Systems-Level Analysis for Understanding Protease Network Operation

Abstract

Extracellular proteases play critical roles in extracellular matrix remodeling and cell migration, both in normal physiology and in diseases such as cancer. However, complex networks of protease crosstalk, substrate regulation, and feedback interactions have limited our ability to understand and therapeutically target these enzymes. Recent developments in experimental methods allow quantitative characterization of multiple pathways within the protease network. However, mathematical modeling has yet to be appreciably utilized for interpreting relationships between proteases, regulatory processes that govern their activity, and affected phenotypes such as cell migration. In this chapter we offer the cue-signal-response paradigm as an especially promising approach to parsing complex biochemical interactions in the protease network, and suggest experimental and computational methods appropriate for such data-driven investigations.

1.1 Introduction

Proteolytic remodeling of extracellular matrix (ECM) plays a central role in physiological processes such as embryological development and wound healing, and is implicated in diseases including arthritis, asthma, and cancer (Overall and Blobel, 2007). Matrix metalloproteinases (MMPs) have been perhaps the most well studied ECM-degrading enzyme family, particularly in the context of mediating cellular migration and cancer metastasis (Kessenbrock et al., 2010). However, research over the last decade combined with discouraging clinical trial results for MMP inhibitors have shown that metalloproteinases (MPs), and indeed many extracellular protease families, exhibit multifaceted and context-dependent behavior (Kessenbrock et al., 2010; Mason and Joyce, 2011). Many MMPs degrade cell-signaling molecules, protease inhibitors, and proteases themselves, in addition their canonical ECM substrates. The complex set of biochemical interactions among proteases, their substrates, and their regulating pathways constitutes a "protease network" not easily interpreted on a component-by-component basis.

This thesis chapter offers the cue-signal-response (CSR) paradigm as a multivariate, quantitative, systems-level approach to understanding the protease network, especially as it relates to cell migration. Major progress has been made in developing various global profiling, or "omic", methods to catalog the numerous biochemical interactions among proteases and their interacting partners (Impens et al., 2010). Nonetheless, opportunity remains to use this catalog of biochemical interactions to build quantitative, data-driven, and predictive models of protease network behavior in development and disease. The feedback, cross-talk, and cooperative non-linear behaviors exhibited by protease networks compel the application of systems-level conceptual frameworks and computational methodologies that have been successful in other similarly challenging applications, including intracellular phosphosignaling networks (Prill et al., 2011; Hyduke and Palsson, 2010; Morris et al. 2010), metabolic circuits (Oberhardt et al., 2009) and the integrated modeling of an entire organism (Karr et al., 2012) or disease process (Akavia et al., 2010).

In this essay, we begin by briefly defining the landscape of key extracellular proteases related to cell migration, and provide recent examples of how protease network complexity has produced abstruse results. We describe the CSR framework for integrating computational methods and experimental tools to understand protease network function, and provide examples of its successful application. Finally, we describe how CSR modeling could aid in the development of therapeutic strategies that target the protease network.

1.2 Protease Network Components and Interactions

Recent advances in genomic and proteomic methods, combined with a growing appreciation for the multiple roles of proteases in vivo, have illuminated the extent of complexity in extracellular protease biology in recent years (Rodrguez et al., 2010) Fig. 1-1. Much of this information has been cataloged in easily accessible formats, such as with the MEROPs database (Rawlings et al., 2012). Of the over 250 extracellular and cell-surface proteases in the human genome, three enzyme families have received the bulk of attention relating to ECM degradation and cellular migration: the 23 members of the MMPs; the 8 members of the related A Disintegrin and Metalloproteinases (ADAMs) that are somatically expressed and catalytically active (Primakoff and Myles, 2000); and the roughly 15 cathepsins, which comprise serine, aspartyl, and cysteine proteases known for lysosomal proteolysis at low pH (Overall and Blobel, 2007). Members within all three groups directly proteolyze ECM components. Other relevant proteases include tissue- and urokinase- plasminogen activators (tPA and uPA, respectively), which degrade ECM components chiefly through the activity of plasmin. The tolloid family of proteases, including bone-morphogenic-protein-1 (BMP-1), also degrade ECM components and are especially involved in cartilage and bone development. Closely related to ADAMs, ADAMs with thrombospondin motifs (ADAM-TS) proteases are a large family of enzymes that most

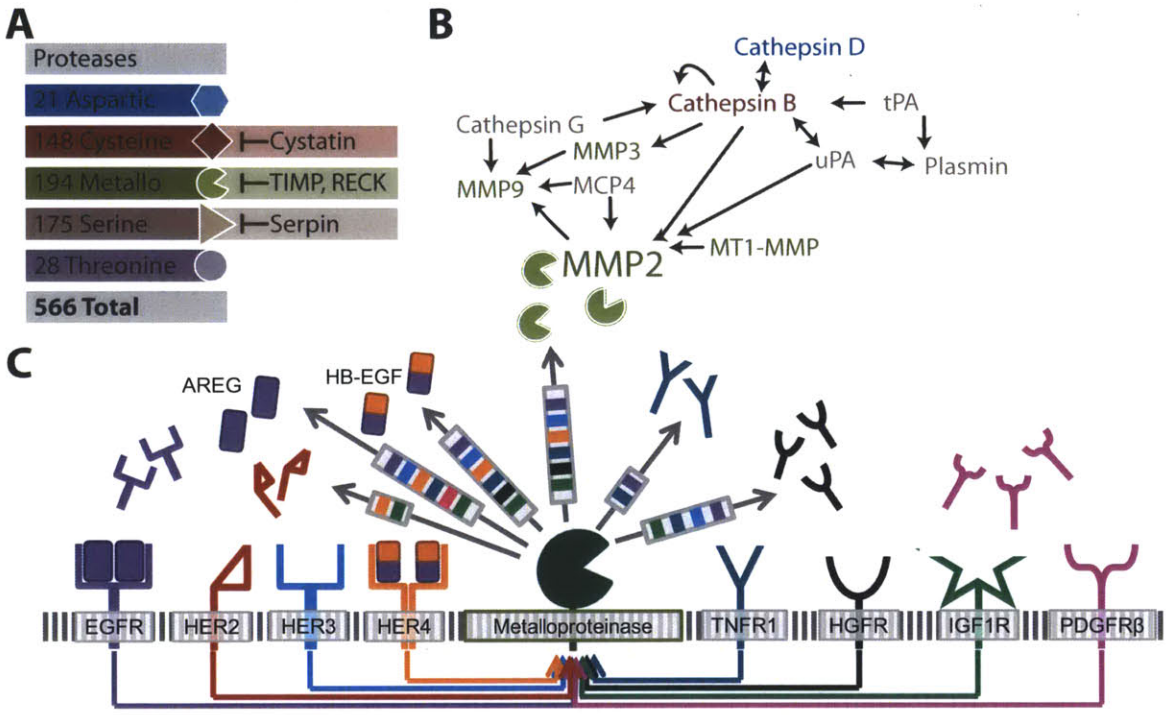

Figure 1-1: **Overiew of the protease network.** A) Proteases of the human genome, defined by active site chemistry, and their inhibitors. B) Proteases are enmeshed in a complex and recursive network of biochemical interactions. Arrows indicate a known proteolytic cleavage reaction, and colors correspond to active-site chemistry depicted in *A*. Interaction network is not exhaustive, but nevertheless conveys feedback relationship among closely interacting enzymes. C) More thoroughly studied proteases such as ADAM-10 and -17 have revealed a dense network of both upstream signaling regulation and promiscuous interaction with substrates and other proteases. In Chapter 4 of this thesis, all cell-surface receptors depicted here were stimulated with exogenous ligand and found to modulate supernatant accumulation of various substrates in a MP-dependent manner (shown by arrows; bars on arrows indicate supernatant levels change following receptor stimulation, corresponding by color).

prominently degrade glycoproteins and proteoglycans such as aggrecan. Characterization of newly uncovered protease families continues to be an active area of research. Type-II transmembrane serine proteases (TTSPs), including matriptase and hepsin, cleave diverse substrates including uPA (Bugge et al., 2009). Meprin-beta is another recently characterized MP capable of cleaving and activating ADAM-10 (Jefferson et al., 2013). Although not an exhaustive list, the roughly 50 enzymes discussed above constitute a majority of actively investigated proteases involved in cellular migration.

These key proteases participate in a dense network of biochemical interactions. Numerous mechanisms of feedback, cross-talk, and cooperativity that characterize the protease network have been extensively described in several recent reviews (Kessenbrock et al., 2010; Mason and Joyce, 2011; Duffy et al., 2011). MMP, cathepsin, and ADAM(TS) family enzymes are generally capable of cleaving multiple distinct substrates, often with non-exclusive substrate preferences. Proteomic studies have documented these substrates, which include structural ECM components such as collagen; growth factor and cytokine ligands and receptors; adhesion ligands and receptors; endogenous protease inhibitors; and extracellular proteases themselves (Rodrguez et al., 2010; Impens et al., 2010). Proteolysis generally modifies substrate bioactivity in complex manners, for example by modifying ECM binding sites on degraded chemokines (Starr et al., 2012). Ultimately the complex web of protease-substrate interactions has impeded our ability to interpret the overall impact of individual proteases.

Protease activity regulation is also complex. Hundreds of endogenous inhibitors act upon multiple proteases (Rawlings, 2010). For instance, tissue inhibitors of MPs (TIMPs) block MMP and ADAM activity; serpin serine protease inhibitors block plasminogen activators; cystatins inhibit cysteine proteases including some cathepsins; and Reversion-inducing-cysteine-rich protein with Kazal motifs (RECK) inhibits MMP-9 (Clark et al., 2007). Additional protease regulation occurs through proteolytic processing and trafficking, for example by the rhomboid family member iRhom2 (Adrain et al., 2012; McIlwain et al., 2012), along with ectodomain shedding of the mature protease from the cell surface. ADAM-10 activity, for instance, can be regu-

lated by ADAM-9 shedding it from the cell surface (Moss et al., 2011), and ADAM-9 can in turn be shed by MT1-MMP (Chan et al., 2012). Cell surface protease activity responds to intracellular signaling pathways, although the exact mechanisms remain uncertain. For example, regulated dimerization (Xu et al., 2012) and phosphorylation of the C-terminus (Xu and Derynck, 2010) have been described as a possible mechanism of controlling ADAM-17, yet other reports demonstrate induced ADAM-17 activity in a manner independent of the C-terminus (Hall and Blobel, 2012). C-terminal phosphorylation also regulates MT1-MMP through control of its endosomal trafficking (Sugiyama et al., 2010). In addition to regulation of the protease itself, several reports have documented substrate-specific pathways of proteolytic regulation. For instance, specific growth-factor ligand shedding events can be regulated by distinct PKC isoforms (Dang et al., 2011; Kveiborg et al., 2011). ECM components and mechanical forces regulate cell-surface MMPs and ADAMs through integrin engagement, especially relevant to arthritis (Gooz et al., 2012; Leong et al., 2011; Mori et al., 2013). Overall, a clear theme of feedback control has emerged: extracellular proteases cleave the very substrates that govern their behavior, including signaling molecules, ECM components, proteases, and endogenous inhibitors.

1.3 Manifestation of Protease Network Complexity

Complexity in the protease web has frustrated efforts to understand and clinically target extracellular proteases. As evidence, failed clinical trials of first-generation MP inhibitors actually worsened outcomes in some cases. These failures were largely attributed to poor drug specificity, and recent research has shown that some proteases, for example MMP-8, represent anti-targets in several contexts (Decock et al., 2011). Subsequent drugs have been optimized for selectivity. However, even perfectly selective inhibitors will face problems owing to underlying cross-talk and feedback. For instance, genetic deletion of MMP-2 leads to global in vivo changes in protease

expression and activity, clouding the biological implications of MMP-2 as a drug target (auf dem Keller et al., 2013). As another example, ADAM-10 continues to be pursued as a target in breast cancer, chiefly for its role in shedding mitogenic epidermal growth factor (EGF) ligands from the cell surface. Although results show that ADAM-10 inhibition blocks growth factor shedding and tumor growth in several contexts (Fridman et al., 2007; Witters et al., 2008), other reports show inhibition actually enhances metastasis through reduced shedding of the ADAM-10 substrate and proto-oncogene c-MET (Schelter et al., 2011).

The role of proteases during migration through ECM has become an area of focus coinciding with a shift of experiments from 2D systems to more relevant 3D models (Cukierman et al., 2001; Fraley et al., 2010; Meyer et al., 2012). Unfortunately, targeting a single migration modality can be met with compensatory mechanisms (Petrie et al., 2012; Wolf et al., 2003; Guiet et al., 2011). For instance, in some contexts cells can switch from a mesenchymal-like to amoeboid mode of migration in response to broad-spectrum protease inhibition (Wolf et al., 2003; Sabeh et al., 2009). Overlapping substrates among proteases suggests that specific protease inhibitors may lead to compensatory activation of alternative enzymes. Redundancies have been observed in studies involving knockout mice, and as a consequence, significant investment has been made into the development of mouse models lacking combinations of two or even three proteases (Gill et al., 2010). Given the large number of possible protease combinations at hand, such efforts - and their therapeutic counterparts - would certainly benefit from a quantitative, systems-level comprehension of overall protease network behavior to inform experimentation and therapeutic design.

1.4 The CSR Approach

To organize multivariate, quantitative studies of protease networks, here we offer the CSR abstraction as a flexible and simple paradigm (Hughes-Alford and Lauffenburger, 2012; Miller-Jensen et al., 2007; Lee et al., 2012; Kemp et al., 2007) (Fig. 1-2). The approach simplifies complex networks of molecular interactions into a hi-

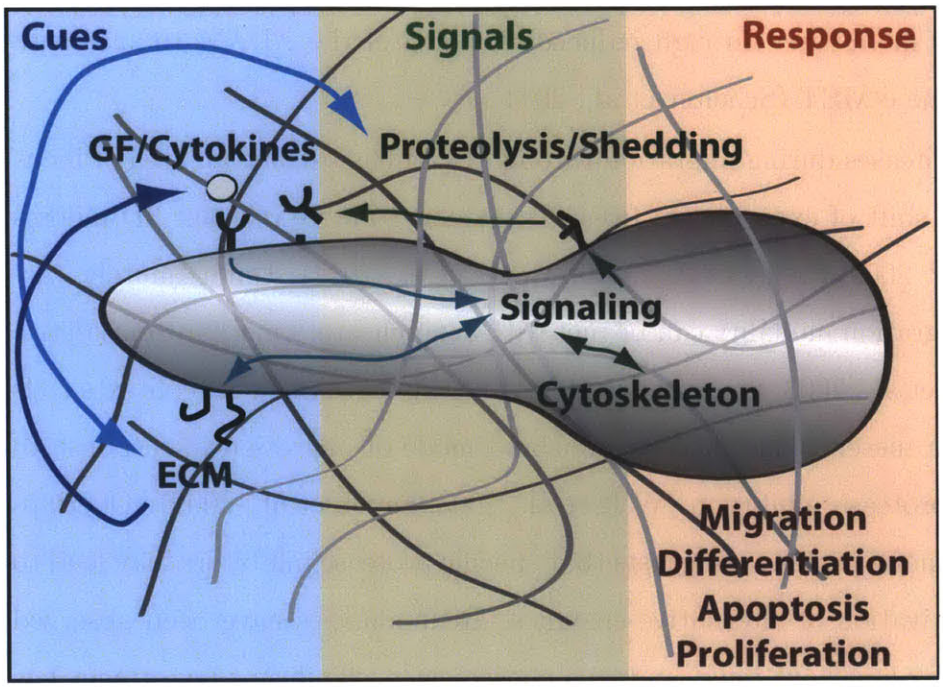

Figure 1-2: **A cue-signal-response paradigm provides a conceptual framework for systems-level modeling.** A cue-signal-response paradigm provides a conceptual framework for systems-level modeling. "Cues" encompass exogenously manipulated variables, and most often constitute growth factor or cytokine treatment, or manipulation of the ECM environment. Measured "signals" often include substrate shedding, direct protease activity, post translational modifications, or cytoskeletal features. "Responses" are those measurements of interest for prediction, typically including cell phenotypes such as migration, differentiation, apoptosis, or proliferation, but can also be protease activities or other measurable quantities.

erarchy where molecular or cellular "cues" affect a cascade of biochemical activities, or "signals", which then affect downstream "responses" at the molecular, cellular, or even physiological level, depending on the application. "Cues" usually represent experimental treatments chosen to address a particular biological question, and may encompass microenvironmental features such as ECM composition, growth factors and cytokines, genetic manipulations, or pharmacological treatments. "Signals," by definition, involve downstream molecular features involved in processing upstream cues. For example, in studies that examine phosphosignaling networks, the "signals" generally comprise concentrations of cell-signaling proteins themselves, such as phosphorylated kinases and transcription factors (Lee et al., 2012). "Signals" are selected to obtain a broad picture of multiple regulatory pathways, especially those which are therapeutically relevant (Duncan et al., 2012; Lau et al., 2011). In protease applications, "signals" may encompass features of protease activity, such as substrate cleavage rates or the active protease concentrations. Finally, "responses" relate signals and cues to a downstream phenotype, and typically include cell processes such as apoptosis, proliferation, and cell migration (Lee et al., 2012; Lau et al., 2011; Kim et al., 2011; Platt et al., 2009). Broadly speaking, definitions of cues, signals, and responses are application-specific: cues in one application may be better defined as signals or responses in the next, and vice versa (Park et al., 2012; Kleiman et al., 2010; Morris et al., 2011; Huang and Fraenkel, 2009). Although other integrative modeling techniques have been reviewed elsewhere, particularly in the context of cancer (Pe'er and Hacohen, 2011), the CSR framework is particularly valuable for understanding and targeting dynamic networks of biochemical activities. The next sections of this chapter describe three components required for successful CSR implementation: manipulation of the network into distinct states through a sufficiently diverse set of cues, measurement of appropriate signals and responses downstream of the cues, and computationally modeling of the assembled data for biological interpretation and prediction.

Measure: Direct measurement of specific protease activities, rates of cleavage for specific substrates, or ideally both in parallel is critical to understanding protease network function. Due to complex regulation, indirect markers of activity such as protein concentration are often uninformative, especially in the context of rapid dynamics. Ubiquitous enzymatic cross-talk, cooperativity, and feedback within protease networks generally require the simultaneous measurement of multiple activities to fully understand systemic patterns of network regulation. Finally, the high degree of interconnected biochemical interactions compels a diverse and numerous set of "cues" by which to observe the network. One strategy for elucidating a CSR landscape involves combinatorially stimulating distinct pathways while simultaneously blocking effector molecules within those pathways, consequently revealing causal relationships between stimuli, effector molecules, and downstream phenotypic outcomes (Saez-Rodriguez et al., 2011).

These considerations imply that the most useful technologies for studying the protease network measure activity directly and in a high-throughput and multivariate manner. Unfortunately, few techniques achieve all three abilities (Fig. 1-3), and broadly fall into one of three categories: singleplexed and having the ability to measure only one specific molecular feature, multiplexed and having the ability to simultaneously measure multiple molecular features, and global "omic" approaches that generally identify the most significant molecular features among hundreds or thousands in a complex biological sample.

Global proteomic technologies have been developed for assessing protease activities, and global proteolytic profiling is rapidly becoming sufficiently quantitative for detailed computational modeling. For analysis of substrate cleavage activity in samples, peptide libraries can be used in combination with mass spectrometry to globally gauge the sequence preference of a sample's proteolytic activity (Schilling et al., 2011; Gosalia and Diamond, 2010). TAILS N-termini labeling can identify cleaved substrates both in cells and in vivo, providing a very broad picture of endogenous protease substrates (Kleifeld et al., 2011). To directly measure active proteases rather than their substrates, activity-based probes have additionally been used as affinity

reagents for mass-spectrometry proteomic applications (Deu et al., 2012). However, such discovery-based mass spectrometry methods are time-, cost-, and technically limited in multiplexing to usually fewer than ten samples, and suffer from significant run-to-run variability.

In contrast to global discovery-based methods, several targeted technologies offer a much higher sample throughput combined with the potential for multiplexing. Traditional antibody-based methods can assess protease expression levels in a high-throughput manner using spotted or solution-phase array technology. Solution-phase antibody arrays comprising dye-labeled polystyrene (Luminex Corp.) or optically-patterned hydrogel (Appleyard et al., 2011) microparticles are especially attractive for their high-sensitivity and low sample-requirement. Furthermore, antibodies are increasingly becoming available that specifically target active proteases, for example as done recently with ADAM-17 (Tape et al., 2011). Products of proteolysis can also be measured in a targeted manner, for example using antibodies for targeting neoepitopes exposed only upon cleavage by a specific protease (Takahashi et al., 2012). Additionally, several technologies have been developed to directly assess the active kinetics of specific proteases. For instance, synthetic FRET-based peptides have been developed for high-throughput enzyme kinetics assays (Moss et al., 2009). However, poor specificity of the fluorogenic substrates has traditionally precluded their application to complex biological samples. To overcome this issue, the substrates have been used in conjunction with a preliminary immunopurification (EMD Calbiochem), although this method currently is not amenable to multiplexing. Alternatively, Chapter 2 of this thesis describes protease activity matrix analysis (PrAMA) for live-cell, non-invasive monitoring of multiple specific protease activities (Miller et al., 2011). PrAMA uses panels of FRET-based substrates, potentially analyzed in the presence or absence of specific protease inhibitors, to generate a signature of protease activities for a given biological sample. This signature of substrate cleavage rates is then compared with known enzyme-specificities for each of the FRET-substrates, allowing specific protease activities to be computationally deconvolved from the panel of non-specific FRET-substrate readings (Miller et al., 2011). Recent advances have also

been made in multiplexed zymographic methods, although these approaches have the disadvantage of dissociating non-covalent complexes, for example with protease inhibitors (Chen and Platt, 2011).

Cell-to-cell heterogeneity has been an increasingly appreciated feature to understanding many biological processes, perhaps most significantly in the areas of stem cell biology and cancer metastasis. Currently, the vast majority of protease-focused studies examine cells on a population level, averaging over significant heterogeneity among even genetically identical populations. Extracellular proteases play a central role in driving cellular heterogeneity, particularly through their modulation of intercellular signaling processes and local ECM remodeling. In fact, many stem-cell markers are protease substrates, including CD44. Microwells are one useful technology for isolated analysis of secreted cellular factors, including proteases (Yalcin et al., 2012). Biomaterials containing fluorogenic protease sensors reveal activity at subcellular resolution, for instance with in situ zymography (Hadler-Olsen et al., 2010). Furthermore, many of the antibody-based methods described previously are amenable to single-cell flow-cytometric and highly multiplexed mass-cytometric (Bodenmiller et al., 2012) analysis. Although the tools to examine the proteolytic network on a single-cell or subcellular scale typically lack multiplexing capability, this continues to be an active area of investigation.

Measurement of phenotypic behavior such as cell migration also requires quantitative and relatively high-throughput methods. The CSR modeling framework supports simultaneous inclusion of multiple phenotypes, and for cell migration these might comprise metrics of cell speed, directional persistence, mode of migration (amoeboid / mesenchymal / collective), or leading edge morphologies such as lamellipodial protrusion dynamics (Kim et al., 2008; Meyer et al., 2012; Hidalgo-Carcedo et al., 2011). Numerous chemotaxis assays have been developed, and are extensively discussed elsewhere (Hughes-Alford et al., 2012). 2D cell cultures allow cell migration and cytoskeleton dynamics to be observed with high-throughput and resolution. Wound healing and plug-closure assays (Gough et al., 2011) enable high-throughput endpoint measurement, while single-cell tracking captures multiple morphological and

migration-related features of heterogeneous cell behavior. Both can be interpreted with automated image-processing packages (Stoter et al., 2012; Sacan et al., 2008; Haass-Koffler et al., 2012; Gobaa et al., 2011) to generate high-content quantitative data. 2D measurements fail to capture many roles of proteolysis, and cell migration depends on ECM composition, mechanics, and geometry (Meyer et al., 2012; Friedl et al., 2012). High-throughput endpoint migration assays quantify migration into a gel from a fixed initial position (such as the bottom of a microtiter plate), and Chapter 4 of this thesis describes an automated software package for performing assays in microtiter format. Single-cell tracking (Kim et al., 2008), organotypic assays that use mixed cell cultures (Gaggioli et al., 2007) and/or decellularized clinical tissue samples as ECM (Ridky et al., 2010), and live-animal models such as the chick chorioallantoic membrane (CAM) model (Ota et al., 2009) provide lower-throughput yet more detailed and physiologically representative descriptions. In rodent models, fine-needle invasion assays and intravital imaging (Entenberg et al., 2013) both have proven useful. Regardless of methodology, the ECM components, molecular organization, and other physiochemical properties should be carefully considered in the context of proteases directly acting upon ECM and adhesion molecules.

Manipulate: Targeted perturbation is essential for CSR modeling (or at the very least, validating results), In addition to cross-talk and feedback issues, proteases often exhibit functions independent of catalytic activity (Mori et al., 2013). Altogether, these effects may lead to inconsistencies between acute versus chronic protease inhibition (Le Gall et al., 2010), or genetic versus pharmacological inhibition (Le Gall et al., 2009). Explicit modeling of these differences within the CSR framework could be an efficient strategy for parsing contradictory results. Within the context of network inference, ideal perturbations are those that stimulate the protease network to identifiably diverse and distinct states. In addition to using combination of genetic/pharmacologic manipulations, this outcome can be realized through environmental stimulations such as cytokine treatment in applications of immunology, for instance. Genetic cDNA and RNAi-based libraries continue to grow more efficient

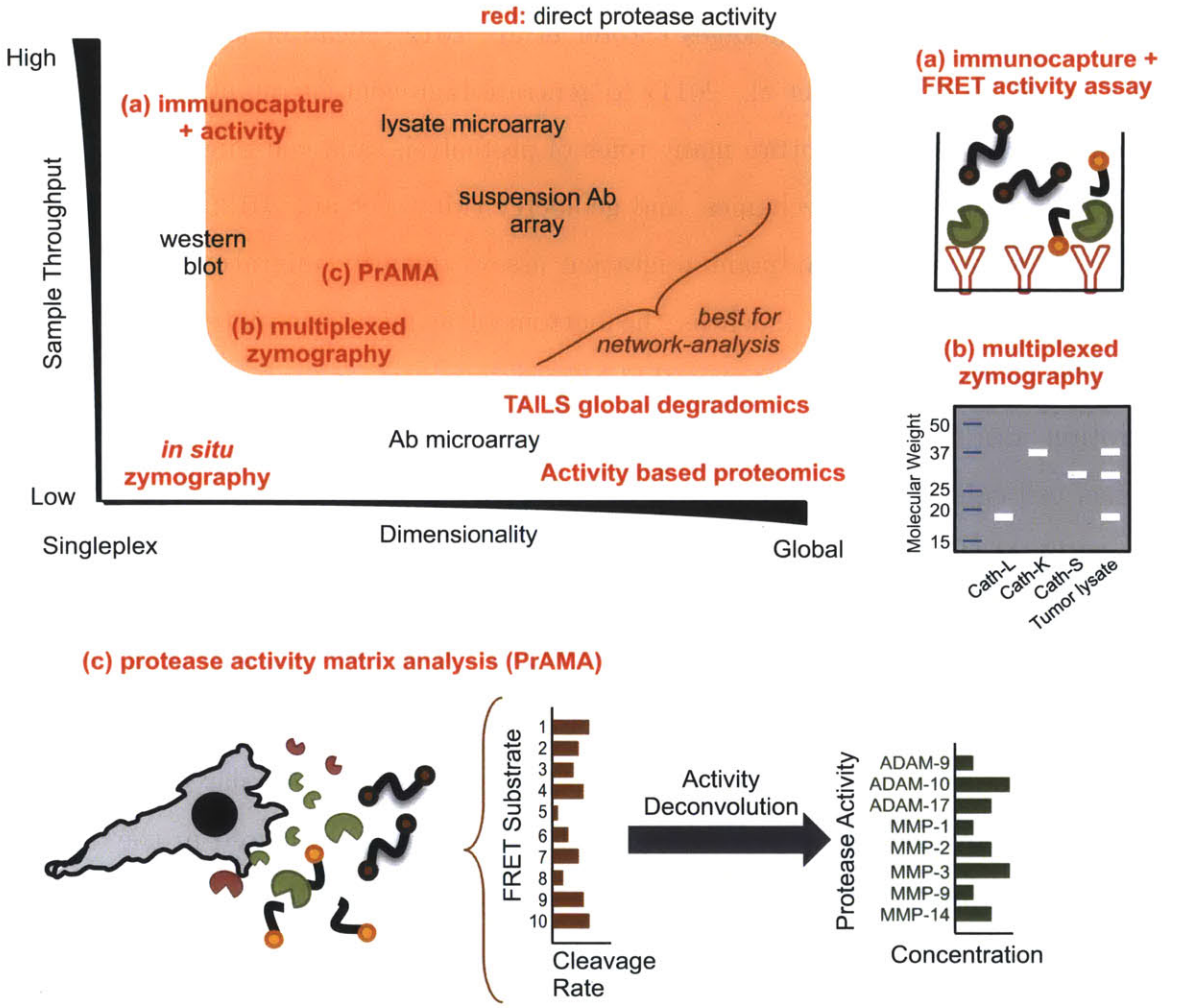

Figure 1-3: **Experimental methods for measuring the protease network.** Network-level investigations into mechanisms of feedback control, combinatorial or cooperative interactions, and dynamic non-linear biochemistry generally require both a high degree of sampling (across multiple time-points, experimental conditions, or perturbations) and a high degree of measurement coverage (across multiple regulatory pathways, molecular features, and phenotypic behaviors). Direct measurement of protease activity and substrate degradation is preferable to assessment of protease expression or secretion. (a) Protease immunocapture followed by FRET-substrate addition can be performed in a relatively high-throughput microtiter format, yet allows for measurement of only one protease activity and removes the enzyme from its in-cell regulatory environment. (b) Multiplexed zymography has been developed as a useful tool for studying cathepsins, in particular, but faces similar issues relating to removing enzymes from their in-cell regulatory partners. (c) This thesis describes PrAMA (see Chapter 2) as a novel method to measure multiple specific MPs in a relatively high-throughput, multiplexed manner in the live-cell context.

and accessible (Pan et al., 2012), and represent a valuable set of tools for specifically manipulating the protease network. These technologies have not been extensively used to specifically address regulatory features of the protease network until very recently, despite their widespread and fruitful application to in vivo and in vitro screens for kinase or integrin activation (Rantala et al., 2011; Meacham et al., 2009).

Model: The successful application of a computational modeling effort is intricately tied to how much information is already known, the scope of data, and the biological question. Modeling techniques range from the data-driven and abstract to the mechanistic and specific, and have been discussed elsewhere, particularly in the context of intracellular signaling networks (Janes and Yaffe, 2006; Morris et al., 2010; Pe'er and Hacohen, 2011). A summary of some methods is presented in Fig. 1-4. Within the CSR framework, more detailed modeling approaches mathematically require either extensive prior knowledge regarding the topology of network-interactions and their rate-constants, or alternatively the largest number of "cues" relative to "signals" to prevent model over-fitting and over-parameterization. Prior knowledge of network topology denotes previous understanding the relationships among cues, signals, and responses, for example as can be gleaned from "interactome" databases such as MEROPS for protease interactions (Rawlings et al., 2012), or using known biochemical rate constants that have been reported in the literature (for instance binding constants between TIMPs and MMPs; Amour et al., 2002). "Omic" scale experiments typically lack the requisite number of "cues" and/or prior knowledge, and therefore are restricted to mechanism-free modeling techniques. Methods such as clustering, principal component analysis, and partial least squares regression aid in reducing data complexity into key clusters or axes of covariation, and can be useful in visualization, classification, and quantitative prediction of overall network behaviors. Statistical models have also been developed to infer network structure, direct biochemical interactions, and even causal network relationships. From frameworks of time-shifted correlation (Bar-Joseph et al., 2012), to mutual-information based approaches (Margolin et al., 2006), to hierarchical multivariate regression tech-

niques (Miller et al., 2012a), statistical network inference has been successfully used to identify novel biochemical interactions and key regulatory processes in various applications.

More detailed modeling techniques can be helpful for studies with extensive prior knowledge or where the number of "cues" being tested outnumber the "signals". Bayesian network analysis incorporates prior knowledge of even a single causal network interaction to infer a globally directional network structure (Yu et al., 2004; Hill et al., 2012). In contrast to simple correlation networks described above, the causal network structure afforded by Bayesian analysis can better describe network pathways and can better guide manipulation of the network into a particular state. Logic-based modeling methods generally require more extensive prior knowledge, for example based on KEGG pathway annotations. They similarly provide a causal network structure, but within a constrained and interpretable set of interactions for understanding how multiple biochemical species integrate to affect one another in a cooperative, interdependent manner (Morris et al., 2013; Wynn et al., 2012). For networks in which the key molecular players have been identified and connectivity has been well established, highly mechanistic models comprised of differential equations can provide a useful, systems-level understanding of how individual biochemical reactions integrate to influence overall network behavior. Time-dependent models based on mass-action kinetics have been demonstrated as useful for elucidating feedback mechanisms and key biochemical reactions within larger networks (Chen et al., 2009), and have been critical to the biophysical understanding of reaction and diffusion processes related to cell signaling (Berezhkovskii et al., 2009) and cell migration (Painter et al., 2010; Eisenberg et al., 2011).

1.5 Demonstrations of Protease Network Modeling

Several recent examples provide encouraging results into how network-level computational modeling can lead to insights into protease and cell-migration biology. At the mechanistic end of the modeling spectrum (Fig. 1-4), mechanics-based differential equation models have described cell migration as a function of contractile forces, material properties, and enzymatic proteolysis. These studies provide explanations for how biphasic relationships may arise between integrin and MMP expression in the context of mechanical forces driving migration (Harianto et al., 2010). In another example, a partial differential equation model explores the relationship between tumor acidity, proteolytic activity, and tumor invasion, finding low pH may prevent tumor invasion in some contexts (Martin et al., 2010). Studies examine the role of material properties and mechanical forces in affecting cell migration, and quantitatively describe experimentally observed behaviors as a function of multiple competing processes relating to cell contraction, material geometry and deformation, and cell adhesion and traction (Pathak et al., 2012; Corin et al., 2010; Zaman et al., 2006; Yamao et al., 2011). Unfortunately, most biomechanics-oriented studies treat protease activity as a single process and do not account for differences among proteases, signaling feedback mechanisms, or post-translational protease modifications. Nonetheless, future work may incorporate recently discovered protease biochemistry in greater detail, and the CSR framework will be useful for organizing the increased complexity.

Work has also demonstrated the value in abstract, data-driven modeling approaches in studying protease biology. One recent study implemented a CSR approach to understand how kinase activities from primary monocytes impacted morphologic and cathepsin-related responses to cytokine induction of differentiation. In this case, computational modeling drew quantitatively predictive relationships between signaling pathway activities and downstream cathepsin activities that were consistent across patient heterogeneity (Park et al., 2012). This work stands out as directly monitoring and modeling the activity of specific proteases, and future directions may focus on

understanding the downstream effects of these activities. In another example, gene expression profiling across multiple breast cancer cell lines was used to describe morphological features (including migration-related phenotypes) of cells growing in 3D ECM cultures (Han et al., 2010). This data-driven approach identified peroxisome proliferator-activated receptor gamma (PPARg) as a significant correlate with stellate cellular morphology, and validating experiments found PPARg inhibitors to be effective in blocking this phenotype. This work stopped short of investigating the adhesion, protease, and signaling-related mechanisms of the inhibitor treatment, which may be interesting for future studies.

Systems-level approaches are commonly used in biological areas more easily amenable to high-throughput and multivariate measurements, such as studies of global gene regulation and intracellular signaling, and recent studies in these areas have produced results centered on extracellular proteases. For instance, one study used correlation networks to analyze the global gene expression patterns of regulatory T-cells, and found urokinase plasminogen activator (PLAU) to be centrally important in mediating suppressor function in those cells (He et al., 2012). This work implemented a modeling strategy that fits within the CSR framework, where global gene expression "signals" were measured across a set of time-point "cues" post-stimulation, and were used to discriminate cells as either effector or regulatory T-cells based on a correlation network modeling strategy. Interestingly, results highlighted significant interactions between PLAU, kinase signaling pathways, and cytokine production. Unfortunately the mechanisms of PLAU catalytic activity were under-explored in this work, and experimental methods described in this thesis chapter show potential for elucidating the protease-related mechanism.

Network-modeling has also been successfully applied to study proteases using in vivo models, as demonstrated by a recent investigation of skin inflammation (auf dem Keller et al., 2012). In this study, mass-spectrometry based measurements of global proteolytic activities were combined with genetic deletion of MMP-2 to show, from a very broad perspective, how ubiquitous and complex proteolytic degradation changes with phorbol ester treatment. This work identified a MMP-2 substrate,

C1-inhibitor, that regulates bradykinin generation and vessel permeability during inflammation. The report demonstrated novel application of global proteolytic and proteomic profiling methods to an in vivo model, and used straight-forward statistical modeling to identify significant biological changes. In the future, these various data-types, including measurements of expression levels, post-translational modification, and prior knowledge of protease-substrate relationships, could be better leveraged into a single CSR framework for a statistically richer, more integrated and interpretable assessment of network regulation.

In Chapter 4 of this thesis, we combine multiple data-types, including phospho-signaling, protease activity, substrate degradation, and cell migration into a unified CSR framework. Using multiple data-driven modeling techniques including Bayesian network inference, correlation networks, and partial-least-squares regression, we construct quantitatively predictive models that describe endometriotic cell migration as a function of sheddase regulation. The CSR approach was successful here in identifying known regulatory relationships, for example between kinases and their canonical substrates, while also providing novel insight into how multiple proteases concomitantly cleave multiple ligands and receptors to affect downstream cell migration.

Overall, computational modeling within the CSR framework has been successful at elucidating protease and cell-migration biology. Improved tools for directly assessing specific protease activities, coupled with enhanced knowledge of underlying biological chemistry, will allow future network-modeling efforts to be more activity-oriented and mechanistic.

1.6 Clinical Consequences and Conclusions

The failure of first-generation protease inhibitors in clinical trials indicates a need to better understand the complexities of the underlying protease network. A widely accepted view is that more specific and potent protease inhibitors will be proportionally more successful in the clinic (Drag and Salvesen, 2010). However, this exact point of view has frequently been espoused in the design of kinase inhibitors, with

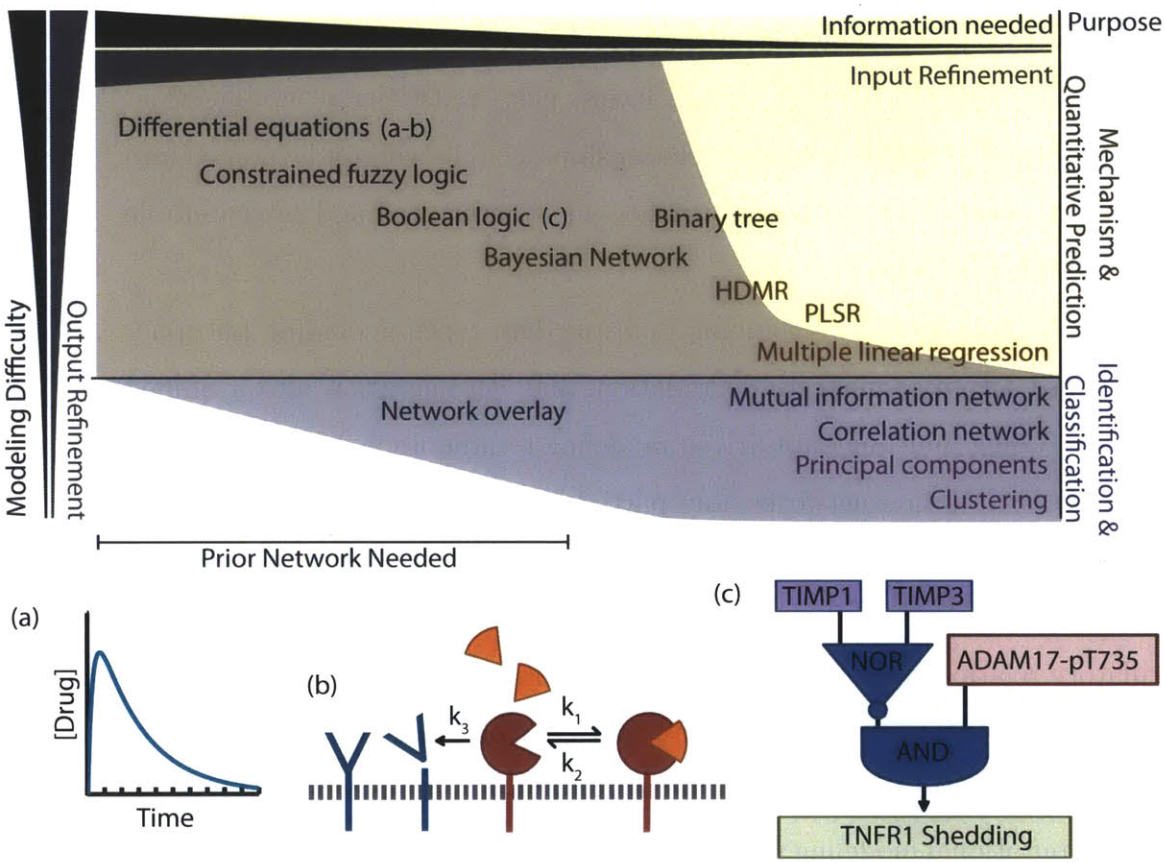

Figure 1-4: **The appropriate computational network modeling approach depends on the type of data, prior knowledge, and intended use.** Larger quantities of appropriately designed, high-quality measurements combined with prior understanding of the system of interest can lend increasingly mechanistic information about the system of interest and a greater predictive capacity. (a) Ordinary and partial differential equations are well suited for pharmacokinetic modeling of protease inhibitor transport at the organismal/physiological levels combined with (b) pharmacodynamic modeling of various competing biochemical processes at the cellular/molecular levels. (c) Logic-based modeling formalisms require less precise prior knowledge, and support inference of combinatorial regulatory interactions. In this toy model, TIMP1 and TIMP3 negatively regulate TNFR1 shedding, which is mediated by active ADAM-17 phosphorylated at T735. HDMR: high dimensional model representation; PLSR: partial least squares regression.

only limited success. Knowledge of overall network dynamics, cross-talk, and subsequent inhibitor resistance mechanisms promises to guide the successful design of multitargeted inhibitors or combination therapies in protease and kinase inhibitors alike (Pujol et al., 2010; Villanueva et al., 2010; Turke et al., 2010). In fact, the close relationships between cell-signaling pathways and extracellular proteases suggest a potential for combination protease- and kinase- inhibitor strategies (Lpez-Otn and Hunter, 2010). Many extracellular protease substrates are already the target of clinically approved therapies, the most prominent example being HER-2. CSR modeling efforts can be useful to identify protease-related mechanisms of drug response and resistance. For instance, computational modeling in Chapter 4 of this thesis describes coordinated ligand and receptor shedding that significantly impacts the migratory response of cells to kinase inhibition. Combination drug treatments are then designed based on this observed mechanism, consequently suppressing compensatory signaling pathways. Overall, computational network modeling promises to be especially useful for designing combination therapies and understanding their mechanism of action.

Interpatient variability has been extensively discussed in the context of kinase inhibitors, and likely plays an equally important role among extracellular proteases as well. Direct measurement of protease activities and their substrates will be valuable for patient stratification and selection into protease inhibitor clinical trials. Protease-activity probe development for in vivo diagnostics remains an active area of investigation, and recent advances include activity-based imaging agents (auf dem Keller et al., 2010), microfluidic-based multiplexed protease activity assays for clinical fluid samples (Chapter 3 of this thesis; Chen et al., 2012), and urinary protease activity markers of disease (Coticchia et al., 2011). Network cross-talk and compensatory feedback mechanisms motivate a multivariate approach to therapeutic and companion diagnostics. Ultimately, a systems-level understanding of basic protease-network biology will be useful to guiding both therapeutic and diagnostic strategies, and the integrative network modeling strategies discussed here aim to assist in that endeavor.

Chapter 2

Proteolytic Activity Matrix Analysis (PrAMA) for Simultaneous Determination of Multiple Protease Activities

Abstract

Matrix metalloproteinases (MMPs) and A Disintegrin and Metalloproteinases (ADAMs) are two related protease families that play key roles in matrix remodeling and growth factor ligand shedding. Directly ascertaining the proteolytic activities of particular MMPs and ADAMs in physiological environments in a non-invasive, real-time, multiplex manner remains a challenge. This work describes Proteolytic Activity Matrix Analysis (PrAMA), an integrated experimental measurement and mathematical analysis framework for simultaneously determining the activities of particular enzymes in complex mixtures of MMPs and ADAMs. The PrAMA method interprets dynamic signals from panels of moderately specific FRET-based polypeptide protease substrates to deduce a profile of specific MMP and ADAM proteolytic activities. Deconvolution of signals from complex mixtures of proteases is accomplished

using prior data on individual MMP/ADAM cleavage signatures for the substrate panel measured with purified enzymes. We first validate PrAMA inference using a compendium of roughly 4000 measurements involving known mixtures of purified enzymes and substrates, and then demonstrate application to the live-cell response of wildtype, ADAM10-/-, and ADAM17-/- fibroblasts to phorbol ester and ionomycin stimulation. Results indicate PrAMA can distinguish closely related enzymes from each other with high accuracy, even in the presence of unknown background proteolytic activity. PrAMA offers a valuable tool for applications ranging from live-cell *in vitro* assays to high-throughput inhibitor screening with complex enzyme mixtures. Moreover, our approach may extend to other families of proteases, such as caspases and cathepsins, that also can lack highly-specific substrates.

2.1 Introduction

Matrix metalloproteinases (MMPs) comprise a family of 23 zinc-dependent endopeptidases that are part of the Metzincin family of enzymes and are generally active on or near the cell surface (Lopez et al., 2007; Kessenbrock et al., 2010). As central regulators of extracellular microenvironments throughout the human body, MMPs play key roles in normal physiological processes including development (Vu et al., 2000), angiogenesis (Rundhaug et al., 2007; Handsley et al., 2005), tissue remodeling (Page et al., 2007), wound repair, and inflammation (Eming et al., 2007). On the other hand, they are also implicated in a wide array of pathologies, ranging from cancer, tumor invasion, and metastasis (Deryugina et al., 2006), to respiratory diseases such as asthma and chronic obstructive pulmonary disease (COPD) (Gueders et al., 2006). MMP proteolytic activities are tightly controlled. Once active, certain MMPs (e.g., MMP2) have been demonstrated to act on hundreds of endogenous substrates (Dean et al., 2007; Morrison et al., 2009). MMP substrates include signaling molecules (e.g., cytokines, chemokines, growth factors, GCPRs, growth factor receptors, and cytokine receptors), extracellular matrix components (e.g., collagen, laminin, and fibronectin), cell adhesion molecules, clotting and complement cascade proteins, and

proteases themselves (Egeblad et al., 2002; Stamenkovic et al., 2003; Myrphy et al., 2008; Blobel et al., 2005).

Closely related to MMPs, ADAM (A Disintigrin and Metalloproteinase) enzymes are metalloproteinases (MPs) within the Metzincin family that are mostly bound at the cell surface (Rocks et al., 2008). At least 13 ADAMs existing in humans have intact metalloproteinase domains and proteolytic activity (Murphy et al., 2008). ADAMs mediate various cellular behaviors including migration, adhesion (Kirfel et al., 2004), proliferation (Blobel et al., 2005), and apoptosis (Duffy et al., 2009; Rocks et al., 2008). Similar to MMPs, ADAM family enzymes are found throughout the body and support diverse physiological processes such as development (Blobel et al., 2005) and angiogenesis (Rocks et al., 2008; Handsley et al., 2005). Likewise, ADAMs can become dysregulated in a variety of diseases and play roles in pathologies including cancer (Duffy et al., 2009; Lopez et al., 2007), inflammatory bowel disease, and asthma (Rocks et al., 2008). Current research suggests that ADAMs have a narrower repertoire of substrates compared to MMPs, and principally function to shed the ectodomain of surface-bound proteins such as growth factor ligands, growth factor receptors, cell adhesion molecules, and cytokine receptors (Murphy et al., 2008; Blobel et al., 2005).

Three key properties of MP biology have created the need for methods that directly observe protease activity in a specific, non-invasive, real-time, and multiplex manner. First, the extensive amount of post-translational modification and regulatory mechanisms controlling MP proteolytic activity make direct activity measurements a valuable and complementary addition to common methods of assessing protein function, such as western blotting, immunohistochemistry, and genetic manipulation (Horiuchi et al., 2007; Xu et al., 2010; Pillinger et al., 2005; Murthy et al., 2008). Second, the plethora of endogenous substrates cleaved by certain MPs, the context dependency of endogenous substrate cleavage, and the overlapping endogenous substrate specificity of closely related MPs make it difficult to quantitatively match the contributions of specific proteases to global observations of endogenous substrate degradation (Horiuchi et al., 2007; Herrlich et al., 2008; Morrison et al., 2009). Quan-

titative determination of selected protease activities would complement measurements that focus on endogenous substrate cleavage, thereby facilitating attempts to match particular proteolytic activities to patterns of substrate degradation. Third, cyclical feedback interactions and compensatory mechanisms among closely related MPs can severely complicate the interpretation of protease network interactions (Joslin et al., 2010; Blobel et al., 2005; Sahin et al., 2004). Non-invasive and multiplexed measurement of MP activity would allow for the assessment of protease network interactions without artificially biasing the underlying network structure.

While many useful methodologies currently exist to study MPs, unfortunately none simultaneously allow for direct, non-invasive, multiplex, real-time measurements of specific protease activity. In general, existing methods such as zymography, activity based probes, and mass-spectrometry based methods all must choose between invasiveness, specificity, and throughput (Butler et al., 2010; Dean et al., 2007; Galis et al., 1995; Kleiner et al., 1994; Sghatelian et al., 2004; Blum et al., 2005). Synthetic polypeptide protease substrates have been extensively developed to directly assess MP activity in a non-invasive and real-time manner (Alvarez et al., 2005; Doedens et al., 2003; Moss and Rasmussen, 2007; Hassemer et al., 2010; Xu et al., 2010). These substrates typically consist of a fluorescence resonance energy transfer (FRET) donor and quencher fluorophore that are separated by a 3–10 amino acid linker containing a protease cleavage motif. Upon cleavage of the polypeptide linker, the donor fluorophore separates from the quencher and fluorescence increases. Protease activity dynamics can then be tracked by observing the change in fluorescence over time. Similarly to many endogenous MP substrates, MP FRET-substrates are generally cleaved by multiple closely related proteases (Moss and Rasmussen, 2007; Caescu et al., 2009; Nagase et al., 1998). Several strategies, including positional scanning of synthetic combinatorial libraries (Lim et al., 2009; Caescu et al., 2009) and directed evolution using phage display (Chen et al., 2002) have attempted to optimize substrate sequences such that they are more selectively cleaved by a specific protease. Combinations of multiple substrates and inhibitors have also been implemented to increase specificity (Rasmussen et al., 2004). These strategies often succeed at dis-

tinguishing between two or a few proteases, but cross-reactivity nevertheless remains problematic in more complex mixtures (Caescu et al., 2009; Drag et al., 2010; Gosalia et al., 2005).

This work describes an approach we term 'Proteolytic Activity Matrix Analysis' (PrAMA) as a method of using panels of FRET-substrates to infer a dynamic, quantitative, and specific profile of MMP and ADAM proteolytic activities. PrAMA ascertains specific protease activity by deconvoluting from measurements derived from relatively non-specific FRET-substrates, employing prior knowledge of individual MMP/ADAM cleavage signatures ascertained using purified enzymes. This approach allows PrAMA to elucidate particular enzyme activities from cleavage signatures obtained in complex samples containing multiple proteases. The integrated experimental measurement and mathematical analysis framework exploits the advantages of FRET-substrates, which support non-invasive real-time measurements of live-cell activity, while addressing their problems of limited specificity and multiplexing. Peptide library microarrays have been previously implemented to assess global patterns of protease activity and infer specific protease activity (Gosalia et al., 2006). Nevertheless, PrAMA's novel combination of mathematical and experimental methodologies allows for greater quantification of protease activity, lower expense, and higher throughput compared to microarray-based approaches. Ultimately, PrAMA fills a niche that complements many other current methods of assaying MP activity and substrate degradation. This niche is especially important for multivariate network-level analysis, where the ability to simultaneously measure multiple MP activities in a non-invasive, real-time, and relatively high-throughput manner confers the greatest benefits. We anticipate that such network-level approaches will be valuable for designing clinical trials focused on MMPs and for illuminating unintended consequences of the many trials that have failed in the last decade (Morrison et al., 2009; Dorman et al., 2010).

We present a compendium on the order of 4000 measurements involving mixtures of FRET-substrates and purified recombinant MPs, and use these measurements to both construct the PrAMA inference parameters and test the limits of PrAMA in-

ference accuracy. *A priori* determination of the PrAMA inference parameters can predict optimal subsets of substrates for distinguishing particular MPs from each other. We demonstrate PrAMA as capable of accurately inferring MP activity even in the presence of background protease activities. Finally, we apply PrAMA to assess the live-cell proteolytic response of wildtype, ADAM10-/-, and ADAM17-/- mouse embryonic fibroblasts (MEFs) to phorbol ester and ionomycin stimulation. Overall, this work presents the foundation, validation, and theoretical analysis of a general methodology that has potential applications ranging from systems biology to *in vitro* inhibitor screening.

2.2 Materials and Methods

Materials — Recombinant human ADAMs 8, 9, 10, and 17 were purchased from R & D systems. The catalytic domains of the following recombinant human enzymes were purchased from Enzo Life Sciences: ADAM12 and MMPs 1, 2, 3, 7, 8, 9, 10, 12, 13, and 14. MMP9 Inhibitor I (Cat. No. 444278) was purchased from Calbiochem. GM6001 was obtained from Enzo Life Sciences. Recombinant human TNFα and EGF were obtained from Millipore (Billerica, MA). 18 FRET-substrates were obtained from BioZyme, Inc. In this work, we refer to substrates as numbers 1–18, and these reference numbers correspond to polypeptide sequences listed in Table A.1. We performed time-lapse fluorimetry using 384-well OptiPlates from Perkin-Elmer and the Spectromax M3 and M2e fluorimeters (Molecular Devices). We used excitation and emission wavelengths of 485nm and 530nm, respectively, for all experiments.

Substrate assays with purified enzymes — For all experiments, substrates were diluted from 5mM stock in dimethyl sulfoxide (DMSO) to a final concentration of 10μM in the appropriate assay buffer. We conducted experiments in four different assay buffers. "ADAM buffer" consists of 20mM Tris, pH 8.0, and 6×10^{-4}% Brij-35. "ADAM buffer" also includes 10mM $CaCl_2$ for experiments involving ADAM8. "MMP buffer" consists of 50mM Tris, pH 7.5, 150mM NaCl, 2mM $CaCl_2$, 5μM $ZnSO_4$, and 0.01% Brij-35. We obtained the third buffer, Clonetics^{TM} Mammary Epithelial Cell Basal Media ("MEBM"), as a phenol-red and serum free solution from Lonza, pH 7.4. Final active concentration of MMPs and ADAMs in activity assays ranged from 0.01nM to 7.5nM.

As a fourth buffer, we spiked purified MMP into the conditioned media from the MDA-MB-231 cell line, which is an estrogen receptor negative (ER–) breast cancer cell line derived from the pleural effusion of a breast cancer patient. We obtained these cells from the American Type Culture Collection (ATCC, Manassas, VA) and routinely cultured them at 37°C, 5% CO_2, in DMEM supplemented with 10% foetal calf serum, 100 U/ml penicillin, 100μg/ml streptomycin, 4mM L-Glutamine, and 4.5g/L

D-glucose. We collected cell supernatant under the following conditions: cells were grown to 80% confluency in 10cm tissue-culture treated polystyrene plates obtained from Corning Life Sciences (Lowell, MA), serum-starved for 4hrs in basal media consisting of DMEM supplemented with penicillin/streptomycin, and stimulated with basal media supplemented with either 10ng/ml TNFα or 10ng/ml epidermal growth factor (EGF). Supernatant was collected at 12hrs, spun down at 200g for 5mins to remove debris, and immediately flash-frozen. For FRET-substrate assays involving this supernatant, final reactions were composed of a 2:1:1 mixture of 20uM substrate diluted from 5mM DMSO stock into phosphate buffered saline, 4nM active MMP7 diluted in "MMP buffer," and thawed supernatant.

We determined active site concentrations by comparing observed cleavage rates to previously published catalytic efficiencies for the same substrates in either "MMP Buffer" or "ADAM buffer" (Moss and Rasmussen, 2007; Mohan et al., 2002). In some cases we performed active site titration with GM6001 to either confirm this comparison or to substitute it when comparison was unavailable. Activity data for active site titrations were fit to the Morrison equation using non-linear least squares curve-fitting (see below). We normalized substrate concentration to a positive control, comprised of 10μM substrate incubated with 0.5% trypsin and 0.2% EDTA (Sigma). Almost all experiments were performed in technical triplicate, except for the MMP7 dilution series and the experiments involving cell supernatant, which both were performed in technical duplicate. For the experiments in triplicate, we excluded clear outliers in a few cases (<10% of all triplicates) using Dixon's Q-test with a 90% threshold. We performed all experiments at 37^{o}C. In general, readings measured fluorescence approximately every half-hour for roughly five hours. We conducted all computational work using Matlab (2009a, MathWorks, Natick, MA).

Live-cell substrate assays — Mouse embryonic fibroblasts (MEF cells) were a kind gift from Drs. Carl Blobel (Hospital for Special Surgery, New York, USA) and Paul Saftig (University of Kiel, Germany). ADAM10 (Hartmann et al., 2002) and ADAM17 (Horiuchi et al., 2007) knockout fibroblasts were derived from E9.5 and

E13.5 embryos, respectively. Cells were maintained at 37°C, 5% CO_2, in DMEM supplemented with 10% foetal calf serum, 100 U/ml penicillin, 100μg/ml streptomycin, 4mM L-Glutamine, and 4.5g/L D-glucose. For PrAMA, we plate 5000 cells per well in clear-bottom 384-well polystyrene plates from Thermo Scientific (roughly 70,000 cells/cm^2). The following day, we change media and add one of seven FRET substrates (substrates 1, 2, 5, 6, 7, 9, and 15) at 10μM to each well, along with either 1μM phorbol 12-myristate 13-acetate (PMA), 10μM ionomycin (IM), or a DMSO control. In all cases media contained <1% DMSO. Following addition of substrate, cells were imaged at 20-40min intervals and at 37°C for 2hrs. We performed experiments in biological quadruplicate and excluded data lying more than 1.5 standard deviations from the mean (at most one sample per quadruplicate).

Enzyme kinetics modeling – We model enzyme kinetics as an extension of the classical Michaelis-Menten (M-M) model, where the initial rate of cleavage, V_0, is assumed to be constant and is defined as the following: $V_0 = C_{i,j}[S_i]_0[E_j]$. $C_{i,j}$ describes the catalytic efficiency, k_{cat}/K_m, with which the jth enzyme $[E_j]$ cleaves the ith substrate $[S_i]$. This model assumes $K_m >> [S]$, which has been experimentally confirmed for several substrates (Moss et al., 2009). We assume minimal inner-filter effect, although this can become significant at substrate concentrations above those used in this work ($> 20\mu$M) (Palmier et al., 2007). In all experiments we aim to infer either the initial rate of substrate cleavage (V_0) or the catalytic efficiency ($C_{i,j}$) in a reaction, depending on whether [E] is known or unknown (i.e., "self-blinded" in this work), respectively. We infer these parameters by fitting a kinetic model to the time-lapse fluorimetry data, where fluorescence $F_p(t)$ indicates product formation. Typically, inference using M-M kinetics involves fitting early time-points to the linear M-M model $F_p(t) = (V_0 t + B)F_0/[S]_0$, where B is the background signal and F_0 is the peak fluorescence from the positive control (described above). We extend the linear M-M description to a non-linear "depletion-decay" model. We let $[S]/[S]_0 = 1 - F_p/F_0$, and describe the observed fluorescence, F_{obs}, in the following equations:

$$dF_{obs}/dt = V_0(F_0 - F_p)/[S]_0 - k_d F_{obs} \quad (2.1)$$

$$dF_p/dt = V_0(F_0 - F_p)/[S]_0 \quad (2.2)$$

Where k_d indicates the first-order photobleaching decay of the fluorescent cleavage product. Eq. 2.2 explicitly accounts for the depletion of substrate reactant as it becomes degraded by active protease, and we discuss substrate depletion further in the supplement Text A.1. Michaelis-Menten kinetics have previously been modeled to account for substrate depletion (Cornish et al., 1995; Robinson et al., 1984; Obach et al., 2002; Nath et al., 2006), and the kinetic model used here varies slightly from previous methods by considering both substrate depletion and photobleaching. We define a lag-time, T_0, that denotes the amount of time between the reaction start and the first fluorimetry measurements; $k_d = 0$ for $t < T_0$. From these equations, the analytical solution of F_{obs} takes the following form:

$$F_{obs}(t) = F_0 V_0 (e^{-V_0 t/[S]_0})(k_d[S]_0 - V_0)^{-1} + Ae^{-k_d t} \quad (2.3)$$

$$A = e^{k_d T_0} F_0 \left[(1 - e^{-V_0 T_0/[S]_0}) - V_0 (e^{-V_0 T_0/[S]_0})(k_d[S]_0 - V_0)^{-1}\right] \quad (2.4)$$

We fit model parameters in several steps. First, we subtract the signal of a negative control (FRET-substrate only) from all other signals. The maximum fluorescence in the positive control, which is generally at the first time point, indicates F_0. We determine k_d from the negative slope of the log-transformed positive control ($k_d = -\frac{d}{dt}\ln(F_{\text{pos. cont.}})$). We obtain the remaining two parameters (V_0 and T_0) by non-linear curve-fitting. In several cases, we explicitly measured T_0 and compared model fitting with and without explicitly defining that parameter. Results indicate that V_0 inference remains consistent regardless of whether T_0 is inferred or measured. In cases where $[E]$ is known (i.e. not blinded), we calculate $C_{i,j}$ by the following relation: $C_{i,j} = V_0/([S_i]_0[E_j])$.

PrAMA Inference— PrAMA inference uses panels of FRET-substrate cleavage measurements, coupled with known catalytic efficiencies, $C_{i,j}$, for all relevant i substrates and j enzymes, to infer specific protease activity from a complex mixture of unknown enzymes. PrAMA operates under the assumption that total observed cleavage for the ith substrate $V_{0,i}$ in a mixture of enzymes is the summation of cleavages from each individual protease in that mixture:

$$V_{0,i} = [S_i]_0 \sum C_{i,j}[E_j] \tag{2.5}$$

In this application initial substrate concentration is equal among all experiments, allowing it to be simplified as a scalar constant. The complete set of catalytic efficiencies $C_{i,j}$ comprises the model's $m \times n$ Jacobian matrix for m total substrates and n total enzymes, divided by the substrate concentration:

$$\mathbf{C} = \begin{bmatrix} \frac{dV_{0,i=1}}{d[E_{j=1}]} & \cdots & \frac{dV_{0,i=1}}{d[E_{j=n}]} \\ \vdots & \ddots & \vdots \\ \frac{dV_{0,i=m}}{d[E_{j=1}]} & \cdots & \frac{dV_{0,i=m}}{d[E_{j=n}]} \end{bmatrix} [S]_0^{-1} \tag{2.6}$$

We combine Eqs. 2.5 & 2.6 to relate the vector of substrate cleavage, $\mathbf{V_0} = [V_{0,i=1}, \cdots, V_{0,i=m}]$ to the catalytic efficiencies $\mathbf{C}$, the initial substrate concentration $[S]_0$, and the vector of specific proteases present in the reaction $\mathbf{E} = [E_1, E_2, \cdots, E_{j=n}]$:

$$\mathbf{V_0}^T = [S]_0 \mathbf{C} \mathbf{E}^T \tag{2.7}$$

We determine $\mathbf{C}$ using mixtures of individual enzymes and substrates as described above. The dimensions of $\mathbf{C}$ depend on the number of substrates used in the experiments and the number of enzymes considered. Both of these parameters can be customized to fit the given application. Once $\mathbf{C}$ has been identified, Eq. 2.7 can readily be solved for $\mathbf{E}$, allowing the activities of specific enzymes ($\mathbf{E}$) to be deconvolved from non-specific cleavage signatures ($\mathbf{V_0}$).

Various methods can be applied to solve Eq. 2.7. In this work, we implement a non-negative least squares algorithm combined with inference sensitivity analysis. We employ sensitivity analysis to quantitatively determine robustness to experimental error and to tune inference sensitivity and specificity. PrAMA inference involves three main procedures once $\mathbf{V_0}$ and $\mathbf{C}$ have been measured. First, we perform a bootstrapping scheme of randomly generating an ensemble of 1000 cleavage vectors $\mathbf{V_0^s}$. Sampling is from a log-normal distribution with $\mu = \mathbf{V_0}$ and a standard deviation representative of the average variance between the experimentally observed and PrAMA expected cleavage rates obtained from PrAMA validation sets of data. Second, we use least squares to solve Eq. 2.7 for every $\mathbf{V_0^s}$ in the sampling ensemble, and compute the mean and standard deviation of the ensemble inference results $\mathbf{E}$. In some cases, we added artificial error to $\mathbf{C}$ for each iteration of the boostrapping scheme, representative of experimentally observed parameter uncertainty. This additional process did not significantly improve PrAMA inference, however, and was excluded unless stated otherwise.

As the third step, we apply a robustness filter to the inference results to tune specificity. This filter, termed the σ_T threshold, roughly defines specific protease activities as significant if they are inferred in more than a certain percentage of the ensemble of inferences. We scale σ_T as a fraction of the inference standard deviation that is subtracted from the mean inference value. For example, setting $\sigma_T = 1.0$ roughly defines protease activity as significant if observed in at least 84% of the ensemble inference results.

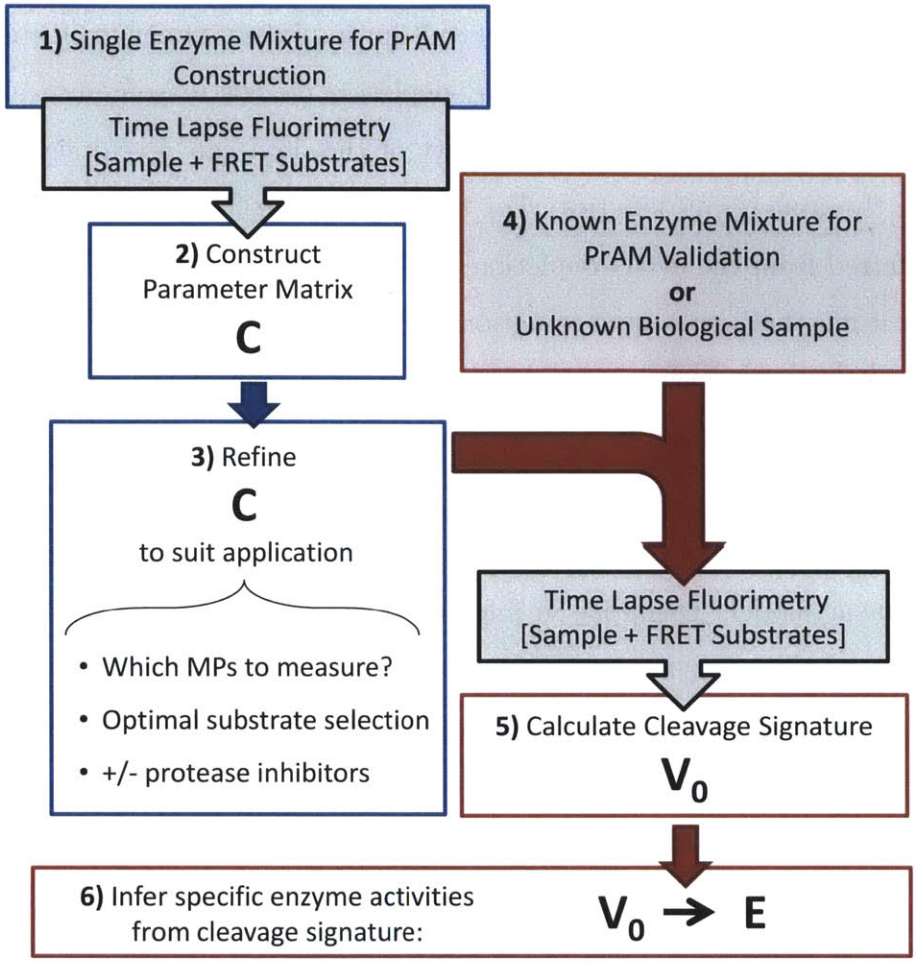

Figure 2-1: **PrAMA overview.** Blue indicates PrAMA development & construction, red indicates PrAMA implementation, and grey indicates experimental preparation & procedure.

2.3 Results

Characterization of the non-linear kinetic model — PrAMA infers specific protease activity levels from panels of kinetic cleavage measurements using FRET-based polypeptide substrates (see Fig. 2-1 for a schematic & illustration of the procedure). The first steps of PrAMA involve determining the kinetic parameters $\mathbf{V_0}$ and $\mathbf{C}$ from time-lapse fluorimetry data (see *Methods*). To establish these two parameters, we employ a "depletion-decay" kinetic model of substrate cleavage that elaborates on linear M-M kinetics to account for first-order photobleaching decay and substrate

depletion. Typical raw time-course fluorimetry output for a mixture of enzymes and substrates can be fit by both the linear and decay-depletion models reasonably well (Fig. 2-2A). However, the often subtle non-linearity of the data can reveal significant differences in the underlying kinetics. For the time-course in Fig. 2-2A, the cleavage rate V_0 inferred from the decay-depletion model is roughly 50%greater than that inferred when using the linear approximation. For a systematic comparison of the two models, we inferred trypsin activity from time-lapse fluorimetry measurements across 2-3 orders of magnitude in enzyme concentration, using both the linear and decay-depletion models (Fig. 2-2B). With log-log scaling, the R^2 coefficient of determination for the decay-depletion model in Fig. 2-2B is 0.9, compared to only $R^2 = 0.25$ for the linear model. Disparity in inference accuracy lies almost entirely at the extremely high and low enzyme concentrations, where substrate depletion and photobleaching, respectively, become most significant. In effect, the decay-depletion model increases the range over which protease activities can be quantitatively and accurately measured by over an order of magnitude.

Several factors make photobleaching a significant issue in this work. We conduct protease activity assays using fluorescein-based FRET substrates over long (often $>$ 5hrs) time scales. Fluorescein is relatively sensitive to photobleaching, and long time scales further amplify photo-sensitivity effects. In this application we typically read fluorescence every 15-30mins and observe resultant first-order decay constants (k_d values) as high as $10^{-4}s^{-1}$. Computational simulations using the decay-depletion model demonstrate how significantly decay can influence the observed fluorescence (Fig. 2-2C). Our results indicate that even the small amount of photobleaching incurred with infrequent plate-reader measurements may result in a several-fold decrease in fluorescence after hours have elapsed.

PrAMA Construction — PrAMA requires the explicit measurements of the entries of matrix $\mathbf{C}$, which characterize the catalytic efficiencies with which individual enzymes cleave FRET-substrates, before specific enzyme activities can be deconvolved from complex reaction mixtures. The dimensions of $\mathbf{C}$ can be customized depending on the experimental application, and Fig. 2-3A depicts one possible con-

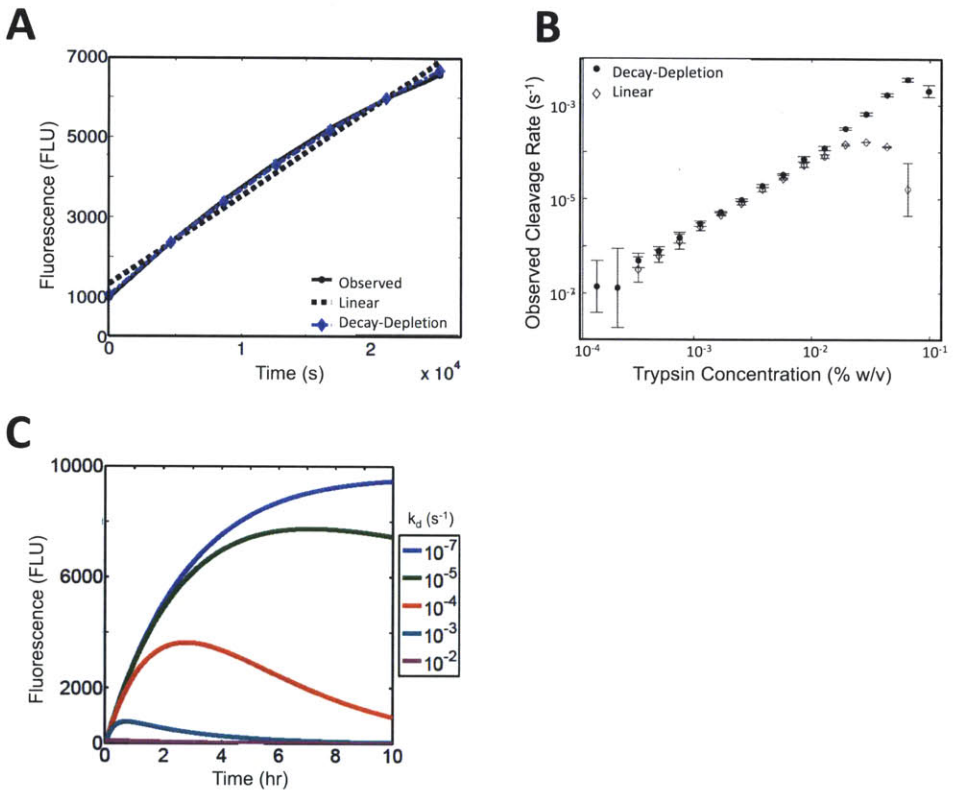

Figure 2-2: **Modeling protease cleavage kinetics.** (A) A typical time-lapse fluorimetry output for a single enzyme (0.5nM MMP10) and substrate 13 in MMP buffer. The data is fit with both linear and decay-depletion kinetic models. The decay-depletion model almost perfectly overlays the data. (B) Inferred kinetic rates ($V_0/[S]_0$) of trypsin cleaving substrate 7 over a range of concentrations. (C) Decay-depletion model simulations of substrate cleavage using the following parameters: F_0 = 10^4 FLU, k_{cat}/K_M = 10^5 $M^{-1}s^{-1}$, [E]=1nM, [S]=10μM.

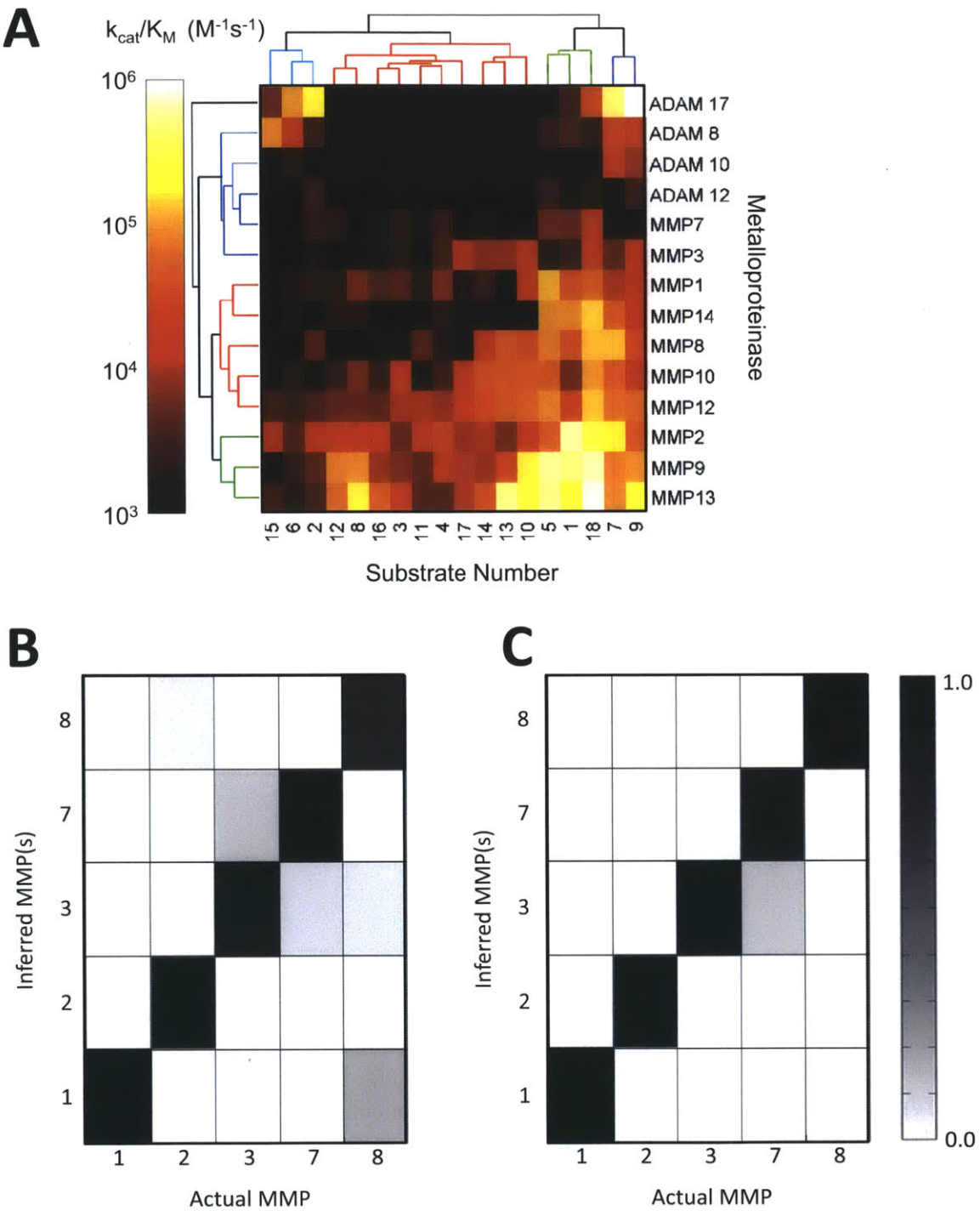

Figure 2-3: **Inference using protease cleavage specificity signatures.** (A) Hierarchical biclustering of observed cleavage efficiencies for various ADAMs and MMPs. Cleavage parameters were averaged over several concentrations. (B-C) PrAMA inference of individual MMPs at different concentrations (B, ~0.5nM vs. ~0.05nM) and buffers (C, MMP minimal media vs. MEBM). Each column represents an individual PrAMA experiment. The abscissa indicates the actual enzyme present, and the ordinate indicates the inferred MMP present. Each PrAMA output (each column) is normalized to have a total signal of 1.

figuration involving 18 FRET-substrates, 4 ADAMs, and 10 MMPs. In this figure, catalytic efficiencies range from $10^3 M^{-1}s^{-1}$ to $10^6 M^{-1}s^{-1}$. Although we successfully measured cleavage rates below this range, such low signals typically have high error and ultimately have low impact on PrAMA inference. We hierarchically biclustered the elements of $\mathbf{C}$ with a Euclidean distance metric and average linkage, without mean-centering or variance normalizing, and with optimal leaf ordering. Several clear patterns emerge from this clustering, as indicated by the dendrograms flanking the array (Fig. 2-3A). MP clustering somewhat recapitulates DNA sequence based phylogenetic profiling of the enzyme families. For example, ADAMs partition from the MMPs, and closely related MMPs such as the gelatinases (MMP2 & MMP9) cluster together. The substrates form three distinct clusters: substrates with greater specificity towards ADAMs (cyan on the dendrogram); substrates with greater specificity towards MMPs (red); and substrates cleavable by both MMPs and ADAMs (blue, green to a lesser extent). More than anything, however, $\mathbf{C}$ explicitly indicates the lack of selectivity among these substrates. Hence, this parameter matrix underscores the need for a deconvolution process in extracting specific protease activities from FRET-substrate measurements, and $\mathbf{C}$ ultimately becomes integral to this task.

A compendium of cleavage signature measurements — We performed roughly 4000 experiments using a variety of enzyme combinations, buffers, and substrates to measure, validate, and test the parameters and principles of PrAMA inference. We conducted a wide array of PrAMA experiments whereby panels of FRET-substrates were used to measure the cleavage signature $\mathbf{V_0}$ of various enzyme mixtures, including roughly 50 single enzyme mixtures, 30 double enzyme mixtures, and 10 triple enzyme mixtures. We tested several buffers, enzyme concentrations, enzyme combinations, and have presented the $\mathbf{V_0}$ values corresponding to these reaction conditions in Fig. 2-4. We hierarchically biclustered the reaction conditions according to their mean-centered and variance-standardized cleavage signatures $\mathbf{V_0}$, using Euclidean distance, average linkage, and optimal leaf ordering. Clustering can often serve as an effective tool to group and/or classify objects. In this application, hierarchical clustering successfully groups over 75% of individual MMPs with themselves, based on their $\mathbf{V_0}$ observed at

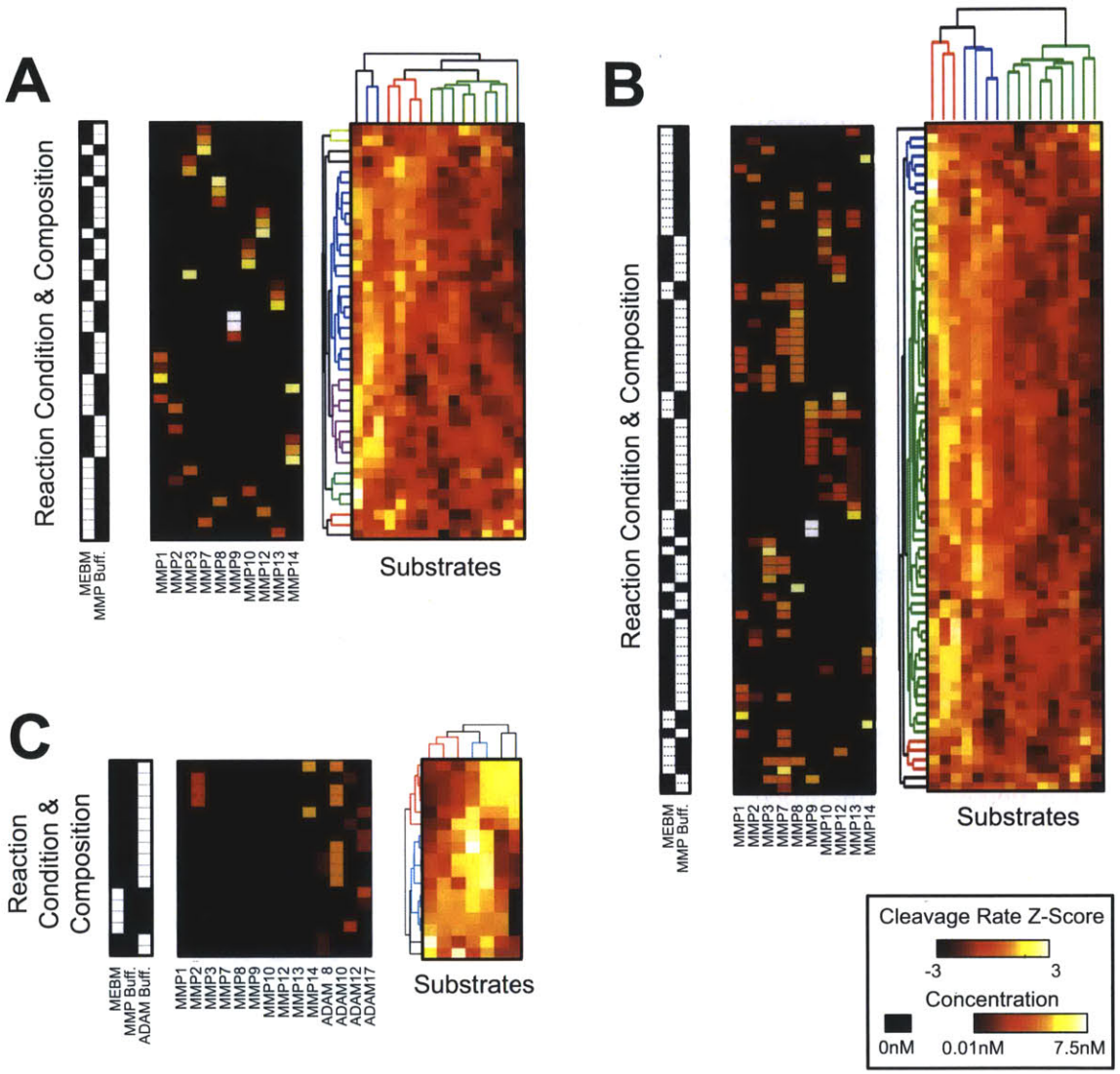

Figure 2-4: **A compendium of cleavage signatures from purified proteases and protease mixtures.** Hierarchical biclustering organizes reaction mixtures by their corresponding cleavage signatures, which consist of cleavage rates (sec^{-1}) of substrates (shown in columns) by various protease mixtures (shown in rows). Within each signature V_0, the Z-score indicates deviation from the mean cleavage rate, following variance-standardization. Three arrays accompany each subplot A-C. The left array describes the buffer (indicated by white) that corresponds to each cleavage signature, aligned by row with the other two arrays. The middle array describes MMP concentrations, and corresponds by row with the adjacent cleavage signatures. Experiment groupings are as follows: (A) all single MMP experiments, (B) all single and mixture MMP experiments, and (C) all single and mixture experiments involving ADAMs.

different concentrations and buffers (Fig. 2-4A). Clustering analysis becomes more difficult to interpret in multi-enzyme mixtures, yet some patterns still emerge (Fig. 2-4B-C). For example, $\mathbf{V_0}$ signatures from mixtures containing ADAM enzymes form two main clusters, corresponding to reactions with and without MMP present (Fig. 2-4C). Nevertheless, simple hierarchical clustering inadequately classifies such complicated enzyme mixtures, motivating a more effective inference methodology to ascertain individual enzyme activities.

PrAMA classifies individual MPs — We first tested the ability of PrAMA to identify an individual enzyme based on its cleavage signature observed at one concentration and/or buffer compared to another. To demonstrate, we constructed a 16×5 parameter matrix $\mathbf{C}$ describing the cleavage of substrates 1–16 by MMPs 1,2,3,7, and 8 at $\{0.4, 0.1, 0.9, 0.7, 1.0\}$nM, respectively, in MMP buffer. We used these parameters to analyze cleavage signatures $\mathbf{V_0}$ from the same enzymes at an order of magnitude lower concentration. Results indicate PrAMA accurately infers the specific MMP based on its cleavage signature, with an average total cross-reactivity of 11% (Fig. 2-3B). As a second demonstration, we used the same parameter matrix $\mathbf{C}$ to infer cleavage signatures corresponding to MMPs 1–8 in MEBM buffer rather than MMP buffer, again at different concentrations. Results for this analysis are even better: 4/5 PrAMA experiments show zero cross-reactivity, and one has 13% cross-reactivity between MMP3 and MMP7 (Fig. 2-3C).

PrAMA inference strengths and weaknesses — We analyzed several properties of the parameter matrix $\mathbf{C}$ in order to *a priori* predict which MMPs PrAMA can accurately infer with high specificity. We transform $\mathbf{C}$ into a model covariance error matrix $\mathbf{R}_M$ that describes inference uncertainty as a function of data uncertainty, $\mathbf{R}_D$, which we directly measure from the variance of replicate experiments. $\mathbf{R}_M$ is mathematically defined as the following:

$$\mathbf{R}_M = (\mathbf{C}^T \mathbf{R}_D \mathbf{C})^{-1} \tag{2.8}$$

The $\mathbf{R}_M$ corresponding to the 16×10 parameter matrix $\mathbf{C}$ characterizing the cleavage

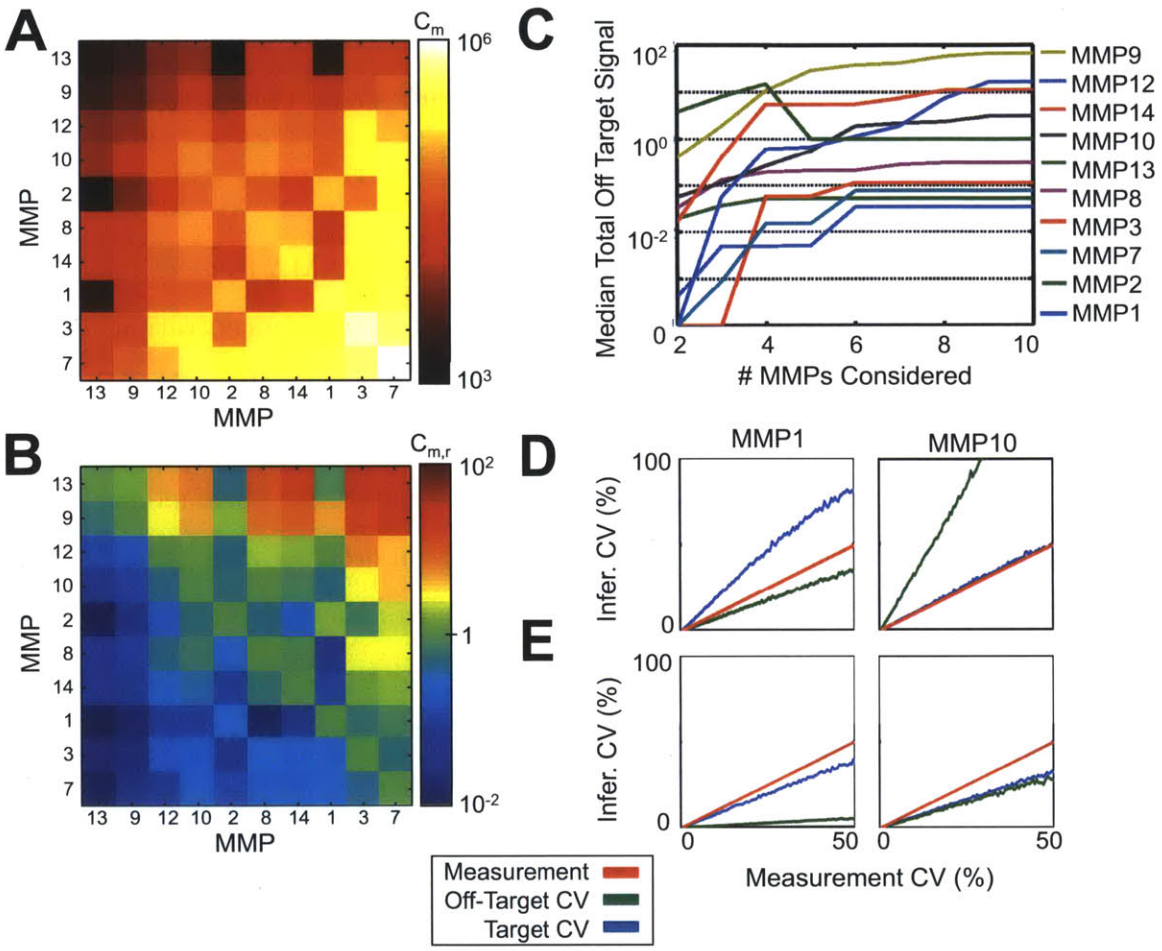

Figure 2-5: **Parameter matrix error analysis.** (A) Model error covariance matrix R_m. (B) Relative model error matrix, R_m^r. (C) Median off-target PrAMA inference error as a function of the number of MMPs considered in the parameter matrix, averaged over all possible combinations of MMP subsets. (D,E) Average target and off-target inference error as functions of the synthetic measurement error, when all MMPs (D) or only MMPs 1 & 10 (E) are considered in the parameter matrix. (C-E) Parameter matrix **C** was constructed using enzymes at ~0.5nM. $\sigma_T = 0$ for all results here. (C) Cleavage signatures were obtained at ~0.05nM.

of substrates 1–16 by 10 MMPs reveals both absolute (Fig. 2-5A) and relative (Fig. 2-5B) amounts of model uncertainty. $\mathbf{R}_M$ can reflect various types of experimental error depending on its construction. $\mathbf{R}_M$, in its form described above, emphasizes multiplicative experimental error and does not *a priori* make expectations regarding the concentrations of particular proteases. Fig. A-1 shows a transformed $\mathbf{R}_M$ that emphasizes additive error, which we experimentally observe to be generally much less significant than multiplicative error.

The diagonal elements of $\mathbf{R}_M$ represent on-target model uncertainties, while off-diagonal elements indicate off-target model error. To emphasize the relative amounts of on- and off- target uncertainty, we subtract the diagonal elements of $\mathbf{R}_M$ from their respective rows to produce a "relative" model uncertainty matrix, $\mathbf{R}_M^r$ (Fig. 2-5B). Large positive values in $\mathbf{R}_M^r$ indicate the potential for high cross-reactivity in PrAMA inference. For example, the $\mathbf{R}_M^r$ rows for MMPs 9, 12, and 13 have large elements corresponding to MMPs 3 and 7. This suggests that signals from MMPs 9, 12, and 13 are likely to be mistakenly inferred as MMPs 3 and 7. We experimentally tested such cross-reactivity by performing PrAMA inference on MMP signals using different $\mathbf{C}$ configurations. We tested all combinations of MMPs considered by $\mathbf{C}$, and performed PrAMA to infer MMP activity from individual enzyme mixtures (Fig. 2-5C). Results indicate that indeed MMPs 9, 12, and 13 have high cross-reactivity with other MMPs. MMPs 1, 2, 3, and 7 are inferred with the greatest specificity. Encouragingly, results suggest that inference cross-reactivity is relatively independent of the number of MMPs considered, past a certain point. For most MMPs, there is hardly any increase in average cross-reactivity when increasing the number of MMPs considered from 6 to 10. For MMP13, the effects of high pairwise uncertainty become diluted when more MMPs are considered, and total cross reactivity actually decreases. To test cross-reactivity as a function of experimental variability, we simulated PrAMA by inferring MMP activity from cleavage signatures generated from the columns of $\mathbf{C}$, but with increasing artificial amounts of multiplicative error added to the simulated cleavage signatures. Fig. 2-5D shows the average results from 1000 iterations of these simulations for MMPs 1 and 10, when all MMPs are considered

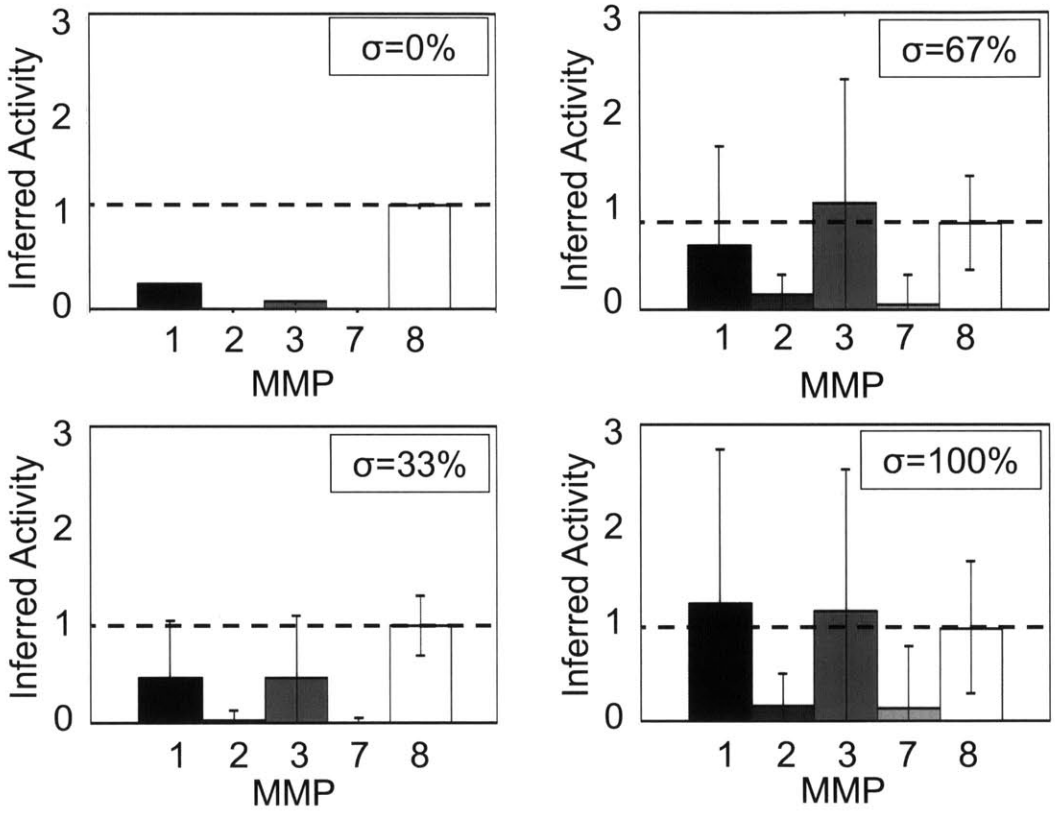

Figure 2-6: **Robustness thresholds filter off-target inference.** Inference results for cleavage signatures obtained using MMP 8 at ~0.05nM and parameters obtained at ~0.5nM. Increasing amounts of multiplicative error was added to cleavage signatures, randomly sampled from a normal distribution with standard deviations of $\sigma = \{33\%, 67\%, 100\%\}$.

in the $\mathbf{C}$ parameter matrix. In agreement with $\mathbf{R}_M$ and $\mathbf{R}_M^r$, MMP1 has higher on-target error, while MMP10 has high cross-reactivity. *A priori* analysis of $\mathbf{R}_M$ can suggest potential protease inhibitors to add or biophysical separation techniques to apply in order to eliminate the number of MMPs considered in a given sample. Both on- and off-target error significantly decrease when MMPs 1 and 10, which according to $\mathbf{R}_M$ have relatively low cross-reactivity, are the only two proteases considered in PrAMA inference. Ultimately this analysis (a) reveals the potential need for additional FRET-substrates with certain specificities, (b) suggests which MMPs can be accurately and simultaneously measured in a given sample, and (c) suggests the potential use of inhibitors or supplementary experimental methods to achieve greater inference resolution. For example, $\mathbf{R}_M$ analysis suggests the potential need for additional substrates that better distinguish MMP9 from MMP7. Although general non-specificity may be difficult to eliminate, substrates could be designed to minimize particular cross-reactivity through a bioinformatic analysis of cleavage motifs (Rawlings et al., 2009; Verspurten et al., 2009) or through targeted directed evolution methods (Ohkubo et al., 2001; Schilling et al., 2007).

Inference sensitivity analysis improves PrAMA accuracy — Experimental error propagates through PrAMA inference in a complex manner. Consequently, we perform a bootstrapping sensitivity analysis to directly account for observed experimental error, gauge its effect on PrAMA inference, and to tune PrAMA specificity/sensitivity (see *Methods*). Experimental replicates of $\mathbf{C}$ and $\mathbf{V_0}$ have an average standard deviation of 20%. Experimental variance is a mix of both additive and multiplicative error, and standard deviation drops to roughly 10% at protease activity levels above $10^4 M^{-1}s^{-1}$. In general, our results suggest that correct PrAMA inference is more robust to experimental and/or artificial noise than incorrectly inferred enzyme activities (i.e. false positive results). As an example, we inferred MMP activity from a cleavage signature $\mathbf{V_0}$ corresponding to a reaction that contained MMP8, but with increasing synthetic multiplicative sampling error applied to the observed V_0 (Fig. 2-6). In this instance, MMP8 does not have the highest average inferred MP activity when synthetic sampling error is high. Nonetheless, PrAMA infers MMP8

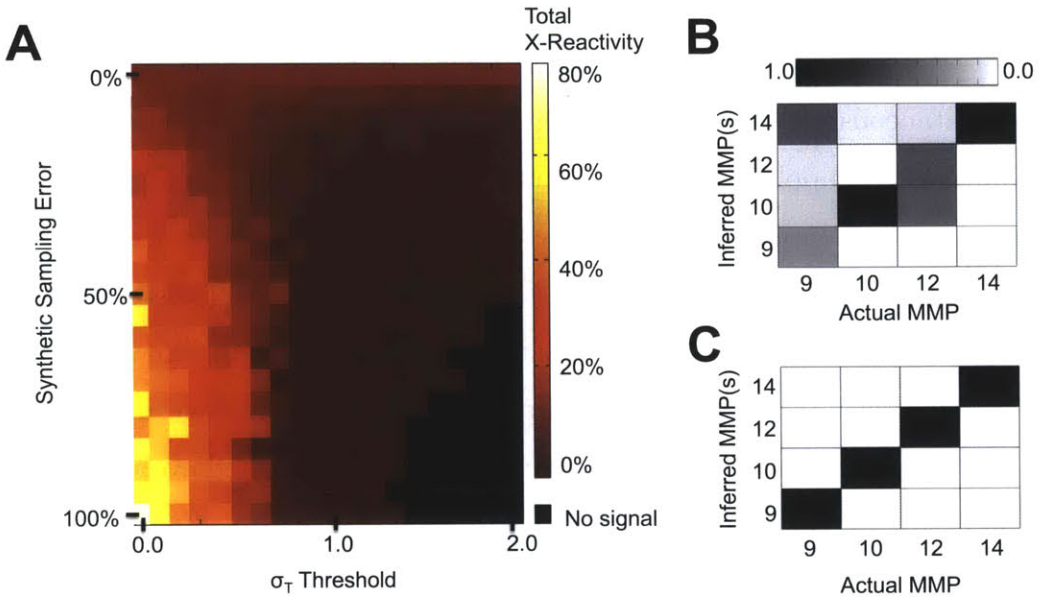

Figure 2-7: **Robustness filter landscape improves inference.** (A) Total cross-reactivity in inference results for cleavage signatures obtained using MMPs 1,2,3,7, and 8 at ~0.5nM and parameters obtained at ~0.05nM. Cross-reactivity is undefined at high sampling error and σ_T thresholds, when no signal falls above σ_T. (B,C) PrAMA inference of MMPs 9,10,12, and 13 at different concentrations (~0.5nM and ~0.05nM) before (B) and after (C) applying 30% sampling error and threshold σ_T=2. Each column represents an individual PrAMA experiment. The abscissa indicates the actual enzyme present, and the ordinate indicates the inferred MMP present. Each PrAMA output (each column) is normalized to have a total signal of 1.

activity with the greatest consistency compared to the other MPs considered. We developed a robustness threshold, termed the σ_T threshold, to take advantage of this general observation. Total cross-reactivity of the inferred MMP activity is a function of (a) the experimental and/or synthetic sampling error of the parameters $\mathbf{V_0}$ and $\mathbf{C}$, and (b) the σ_T threshold. For PrAMA inference of the single-enzyme mixtures involving MMPs 1–8, total cross-reactivity can be totally eliminated by applying the appropriate σ_T threshold, even when the standard deviation of the applied Gaussian multiplicative error approaches 100% (Fig. 2-7A). As another example, the σ_T threshold can completely eliminate as much as 170% cross-reactivity in the inference of individual MMPs 9,10,12, and 13 (Fig. 2-7B-C).

PrAMA inference of MP mixtures We tested the ability of PrAMA to infer specific protease activities from cleavage signatures $\mathbf{V_0}$ observed from 92 total mixtures of one, two, three, and four MPs. We analyzed 48 single-enzyme mixtures, 30 two-enzyme mixtures, and 12 triple-enzyme, and 2 four-enzyme mixtures involving 14 different MPs (see Fig. 2-8A). For these mixtures, we did not consistently observe statistically significant deviation between the observed cleavage signatures V_0 and the expected cleavage patterns based on PrAMA assumptions (e.g., see Eq. 2.7). Figs. A-2 to A-4 show raw time-lapse fluorimetry data, inferred & expected cleavage signatures $\mathbf{V_0}$, and PrAMA inference results for several of these enzyme combinations.

For each mixture, we define MP activity as significant if inferred at levels above a defined σ_T threshold. If that enzyme is actually present in the reaction mixture, then we label the inference for that specific MP as "true positive." Receiver-operator characteristic (ROC) curves then summarize the total PrAMA inference results (Fig. 2-8A). Tuning the σ_T threshold moves inference results along the ROC curve to adjust the true positive and false positive rates. Inference accuracy, defined as the ratio (true positives + true negatives) / (total positives + total negatives), is maximally ~90% for all mixtures, using a parameter matrix that considers the presence of all 14 MPs used in this study. Maximum accuracy for single-enzyme mixtures is slightly above 90%, while more complex triple and four-enzyme mixtures have inference accuracy of roughly 80%. We explored an alternative boostrapping method in this work (see

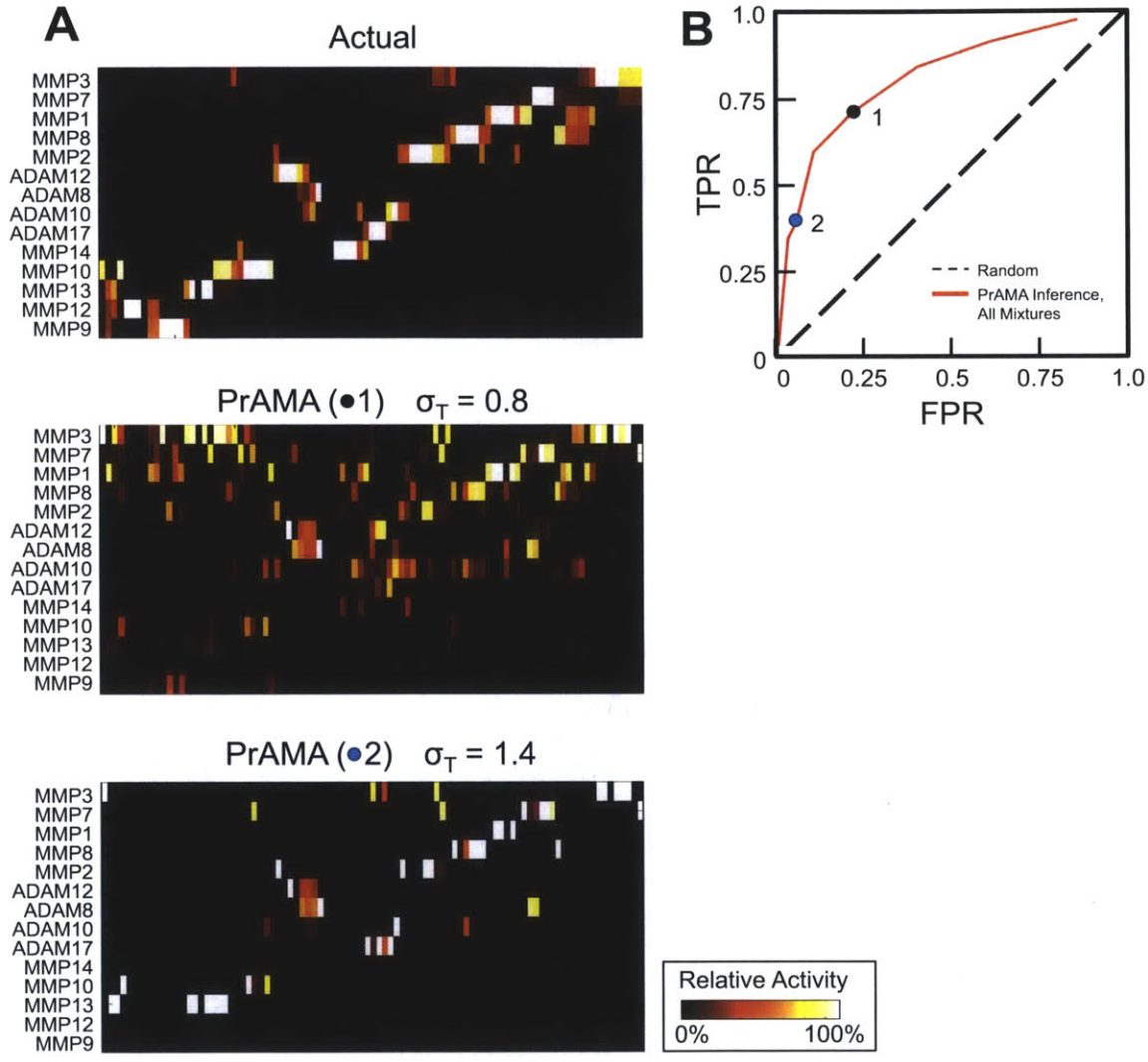

Figure 2-8: **PrAMA inference of enzyme mixtures.** (A) Heat maps indicating PrAMA inference results using mixtures of purified enzymes. Each column corresponds to a different enzyme mixture, and rows indicate which enzymes are considered in the parameter matrix **C**. The top heat map shows actual enzyme activity concentrations, and the bottom two indicate inference results obtained using different σ_T thresholds. Each PrAMA output (each column) is normalized to have a total signal of 1. (B) ROC curve describing PrAMA accuracy for inference shown in *A*. True positive rate (TPR) and false positive rate (FPR) describe accuracies with which PrAMA infers whether or not an enzyme is present in the mixture.

Methods), but found PrAMA results to be robust to the algorithm variation (Fig. A-5). Fig. 2-8A shows PrAMA inference results with increasing levels of σ_T. Previous information regarding MP interactions, co-expression, and cellular localization partly informed the selection of MP mixture compositions. For example, mixtures involving membrane-bound ADAM enzymes also included MMP14, which is the only MMP we analyzed that is membrane-bound with a transmembrane domain. We also included MMP2 in these mixtures, as previous work indicates MMP14 activates MMP2 at the cell surface (Hernandez et al., 2000). We analyzed 5 double-enzyme and 10 single-enzyme mixtures involving ADAM10, ADAM17, MMP2, and MMP14. Fig. 2-9A-B present PrAMA results using a parameter matrix that focuses on just these four enzymes. Maximum PrAMA inference accuracies for both the single and double enzyme mixtures are roughly 90%.

Using PrAMA results from the 92 enzyme mixtures shown in Fig. 2-8A-B, we analyzed which particular proteases PrAMA most reliably measures. We calculated inference accuracy for each of the 14 MPs as a function of the significance threshold σ_T. ROC curves (Fig. A-6) and corresponding maximum accuracies (Fig. 2-9C) for each enzyme reveal ADAM17 to be the most reliably identified MP. Results indicate MMP7 and MMP3 are the most difficult enzymes to measure, and this agrees with previously discussed R_M analysis (Fig. 2-5A). As a caveat, these results may be somewhat skewed by the non-random selection of protease mixture compositions. Nevertheless, results suggest that even enzymes with relatively low catalytic efficiencies for the substrates, such as ADAM12, can be assessed with high accuracy.

We performed PrAMA inference on mixtures containing various concentrations of MMP7 to ascertain PrAMA's ability to quantitatively infer MMP activity, in addition to simply inferring whether or not an enzyme is present (Fig. 2-9D). We used 16 substrates and considered 10 MMPs (i.e., constructed a 16×10 matrix $\mathbf{C}$) in this analysis, and tested 7 concentrations ranging from 0.01nM to 1nM. In all cases, PrAMA inferred MMP7 activity with 100% specificity. Furthermore, PrAMA detected quantitative differences in protease activity with high accuracy. $R^2 = 0.98$ for a log-log plot that describes inferred MMP7 activity as a function of its actual

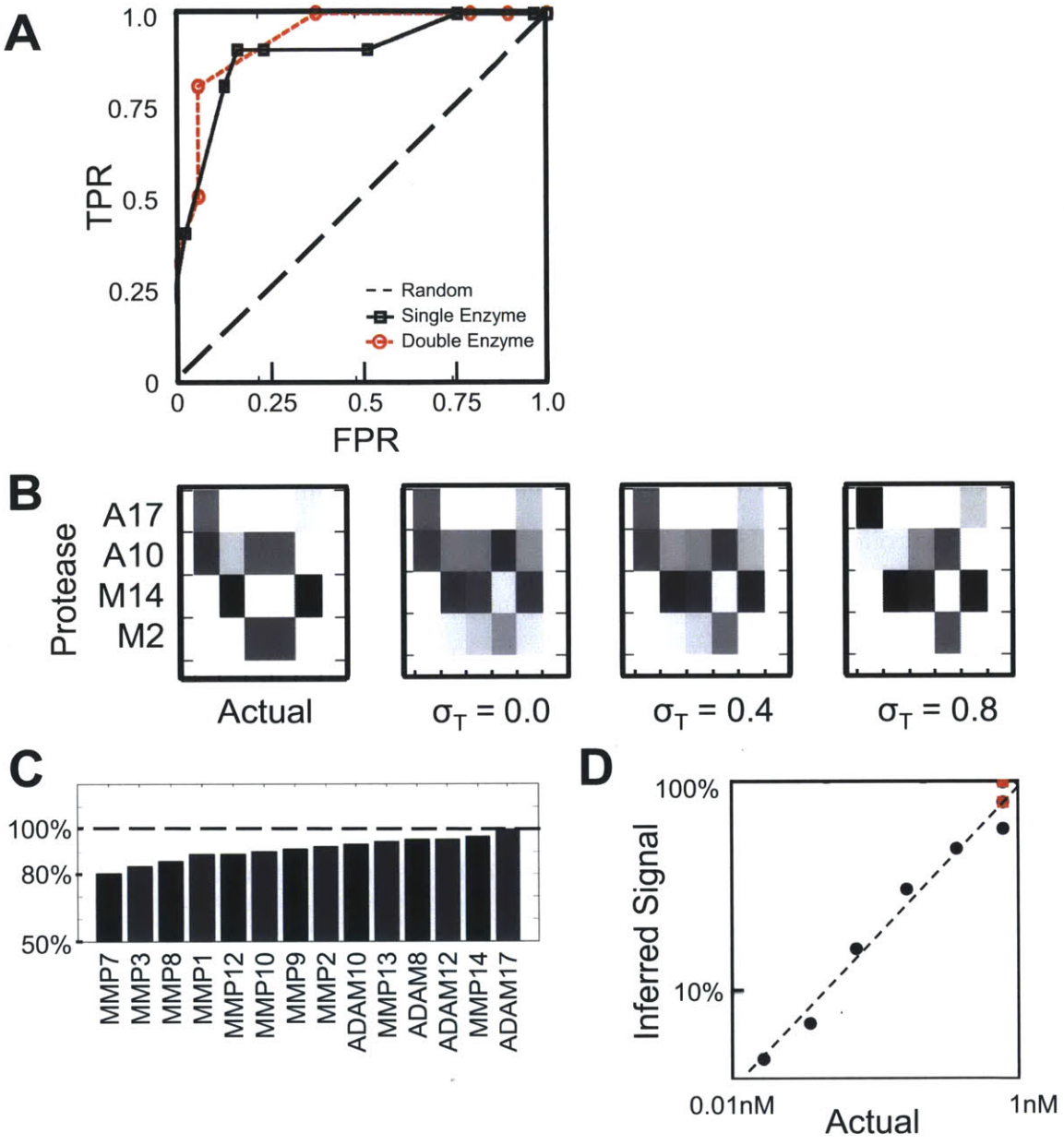

Figure 2-9: **Quantitative PrAMA inference accuracy.** (A–B) Heat maps and ROC curves describing PrAMA inference results for mixtures involving ADAMs 10, 17, and MMP 2, 14. (C) Maximum accuracies for inference of individual enzymes in mixtures shown in Fig. 2-8A. Although 14 enzymes are considered in the parameter matrix as in Fig. 2-8A, here we present inference accuracy for each enzyme individually. (D) PrAMA inference results (ordinate) of MMP7 at different actual concentrations (abscissa). Inference was performed using substrates 1–16 in MMP buffer (black) and conditioned media (red).

concentration. PrAMA has less success in quantifying absolute differences in activity among multiple MPs (Fig. A-7A-B), in part due to the fact that the relationship between MMP concentration and observed protease activity $\mathbf{V_0}$ is enzyme-specific and can deviate from linearity (Fig. A-7C, Fig. A-8). In general, we observe the recombinant MMPs employed in this work to be less efficient at higher concentrations. Enzyme concentration effects on proteolytic activity may be due to issues such as non-specific protein adsorption and aggregation. To test this hypothesis, we added increasing concentrations of Brij 35 to the reaction buffer (Fig. A-9A). Although Brij can decrease proteolytic efficiency, our results suggest that Brij improves assay linearity perhaps by decreasing non-specific aggregation at higher enzyme concentrations (Fig. A-9B), which has been observed for other secreted proteins (Sluzky et al., 1992). Even when nonlinear relationships between MMP concentration and observed protease activity $\mathbf{V_0}$ exist, PrAMA inference does not seem to distort these relationships (Fig. A-7C). Consequently, quantitative comparisons of individual protease activities from one experimental sample to another can still be accurately made.

PrAMA with unknown background protease activity — In many potential applications, PrAMA will not be able to explicitly account for all protease activities in the parameter matrix $\mathbf{C}$. In this work we account for up to 14 MPs simultaneously, and more substrates and enzymes can be potentially included in the PrAMA for future applications. Nevertheless, some biological samples may contain unknown proteinases that are also capable of cleaving the FRET-substrates. Robustness to these unknown proteinases is a crucial property of PrAMA. To test this, we applied PrAMA inference to enzyme mixtures with known "background" protease activity that is not explicitly accounted for in the parameter matrix $\mathbf{C}$. For example, we constructed a 16×2 parameter matrix $\mathbf{C}$ to infer MMP9 & MMP10 activities. We tested PrAMA inference on 5 enzyme mixtures that contained MMP9 and/or MMP10, as well as at least one additional MP that was unaccounted for in $\mathbf{C}$ (Fig. 2-10A). We repeated this process for MMPs 1 8 (3 mixtures) and ADAMs 10 & 17 (3 mixtures). All three sets of analyses performed roughly as well as the PrAMA inference results where all protease activity was explicitly accounted for in $\mathbf{C}$. PrAMA inference of ADAMs

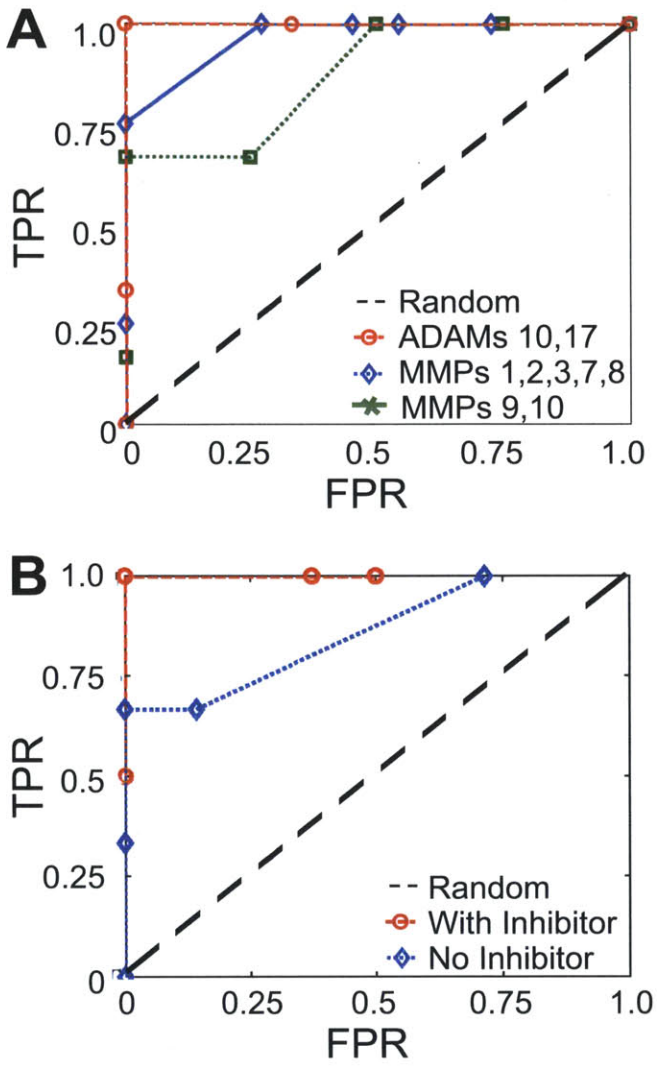

Figure 2-10: **Using PrAMA with background protease activity and protease inhibitors.** (A) The figure legend indicates which MMPs are considered in the parameter matrix. The corresponding ROC curves describe PrAMA inference accuracy for mixtures that contain MMPs and ADAMs that are both considered and ignored in the parameter matrix. (B) Inference for a triple-enzyme mixture involving MMP3, MMP7, and MMP9, with or without invoking information from an MMP9 inhibitor. For inference with inhibitor, $p<0.001$.

10 & 17 yielded a maximum accuracy of 100%, although statistical significance was modest (p=0.17) due to the small sample size (3 mixtures) and low inference dimensionality (only 2 enzymes). Statistical significances of the other two PrAMA results were greater (MMP9 & 10, $p<0.1$; MMPs 1–8, $p<0.001$).

We also used PrAMA to infer protease activity over a background of conditioned media from the breast cancer cell line MDA-MB-231 (Fig. 2-9D). We added recombinant, active MMP7 to supernatant collected 12hrs after stimulating cells with EGF and the inflammatory cytokine TNFα. We considered 10 MMPs in the parameter matrix $\mathbf{C}$, and ultimately were able to identify MMP7 protease activity with 100% specificity. PrAMA did not detect any additional MMP activity in these samples.

Augmenting PrAMA with specific protease inhibitors — The accuracy and specificity of PrAMA can be bolstered by using protease inhibitors in conjunction with traditional PrAMA methods. Adding inhibitors to solutions of active proteases may be appropriate when specificity and accuracy are considered more important than non-invasiveness. In this work, we present an example of how PrAMA can be combined with inhibitors. First, we measured the cleavage profile $\mathbf{V_0}$ of a mixture containing $\sim$ 0.5nM MMP3 & MMP7 and $\sim$ 0.05nM MMP9 in MMP buffer using substrates 1–16. Second, we measured the cleavage rate V_0 of substrate-1 when the mixture had an added 100nM MMP-9 inhibitor (IC_{50} = 5nM). The decrease in observed cleavage rate caused by adding the inhibitor, divided by the previously known catalytic efficiency $C_{i,j}$ and substrate concentration, produced an inferred MMP9 concentration within 20% of the actual concentration: $[E] = V_0/([S]C_{i,j}) \approx 0.04$nM. Based on this, we subtracted the expected MMP9 component of the cleavage signature from the total signature $\mathbf{V_0}$ observed with no inhibitor and performed PrAMA inference on the remaining non-MMP9 component. For this example, PrAMA achieves correct inference of all three enzymes only when incorporating information gleaned from using the inhibitor (Fig. 2-10B). Without this information, PrAMA fails to infer MMP9 without including several false-positives.

PrAMA inference of live-cell response to PMA & IM stimulation — We applied

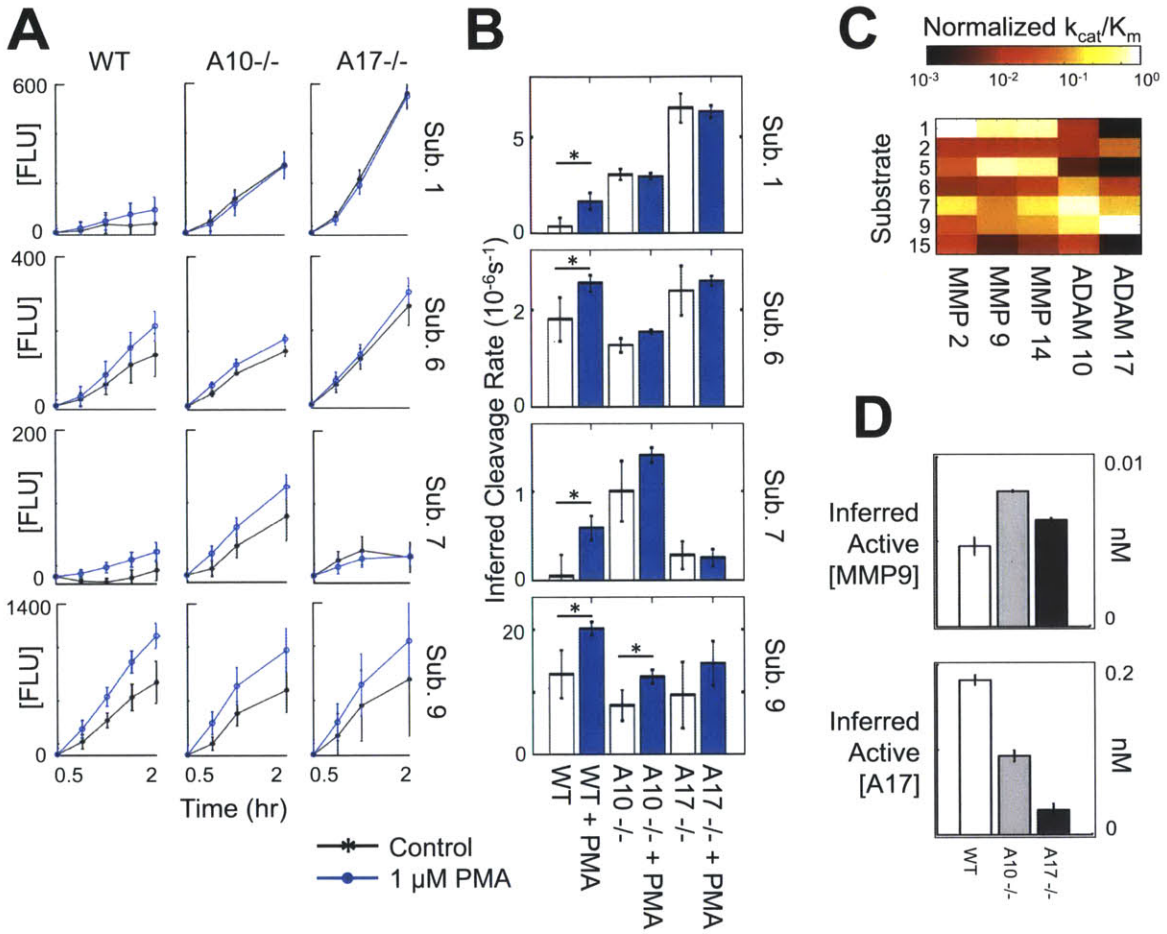

Figure 2-11: **Live cell inference of PMA-stimulated MP activity.** PrAMA was conducted using three cell lines (WT, ADAM10-/-, and ADAM17-/- MEFs) and 7 total substrates, tracking substrate cleavage up to two hours after adding substrate. (A) Time-lapse fluorimetry for 4 of the 7 total substrates used in this experiment. (B) Inferred substrate cleavage rates ($V_0/[S]_0$), corresponding by row to the time-courses shown in A and using all four time-points in A for the inference. Stars indicate $p<0.05$, comparing between cleavage rates for the control and stimulated conditions. (C) Parameter matrix used in this experiment, with each column divided by its Euclidean norm. (D) PrAMA inference results for the increased activity caused by PMA stimulation, using significance threshold σ_T =1.4. No significant increase in MMP2, MMP14, or ADAM10 activity was detected at this threshold. For MMP9 and ADAM17, all inferred differences were statistically significant ($p<0.05$). For all subplots, error bars indicate standard deviation of three biological replicates.

PrAMA to a well-studied set of wildtype, ADAM10-/-, and ADAM17-/- MEF cell lines in order to validate the approach in a live-cell context. Fig. 2-11A depicts the increased substrate cleavage observed in response to PMA stimulation for four of the total seven substrates used in this example. Within 2hrs, PMA causes a statistically significant increase in substrate cleavage for at least one substrate in all three cell lines (Fig. 2-11B): 4/7 substrates significantly increased cleavage with WT cells; 3/7 increased with ADAM10-/- cells; and substrate 15 increased cleavage with ADAM17-/- cells. In some cases, we observe more statistically significant changes in cleavage when including measurements at later time points. Substrates 1, 9, and 15 significantly increase cleavage in ADAM17-/- cells when assessing cleavage 8hrs after PMA stimulation ($p<0.05$). From these measurements it would be difficult to attribute changes in substrate cleavage to particular MPs without PrAMA. We assess protease activity by using a parameter matrix $\mathbf{C}$ that considers the presence of five MPs (Fig. 2-11C). We define the observed cleavage vector V_0 as the change of cleavage rate in response to PMA stimulation, using all seven substrates. In this proof of principle, we limit PrAMA inference to the first 2hrs following cellular stimulation, with the anticipation that later time-point measurements may face greater issues relating to off-target substrate cleavage. PrAMA inference indicates that PMA stimulates significant ADAM17 activity (Fig. 2-11D). In all three cell lines, PrAMA did not detect a significant increase in MMP2, MMP14, or ADAM10 activity. Results indicate that substrate cleavage and subsequent PrAMA are sensitive to treatment with the metalloproteinase inhibitor GM6001 (Fig. A-10). As further validation, PrAMA infers ADAM17-/- cells to have 90% less ADAM17 activity than WT. Remaining ADAM17 signal in ADAM17-/- cells likely arises from other proteases (e.g., ADAM9) with similar substrate selectivity.

We also applied PrAMA to assess the protease-activity response to ionomycin treatment. We used six substrates to ascertain activities of the same five proteases considered in the PMA example, again using wildtype, ADAM10-/-, and ADAM17-/- MEF cells (Fig. 2-12). All substrates show a statistically significant increase in cleavage upon IM stimulation within 2hrs ($p<0.05$), even in the mutant cell lines.

In contrast to PMA stimulation, PrAMA results suggest that IM stimulates MMP9 & ADAM10 activity. In all three cell lines, PrAMA did not detect an increase in MMP2, MMP14, or ADAM17 activity at the same level of significance as detected for ADAM10 and MMP9 (zero at the significance level corresponding to results in Fig. 2-12D). Encouragingly, ADAM10-/- cells show a >90% decrease in inferred ADAM10 activity compared to wildtype cells. Again, remaining ADAM10 signal may be attributed to other proteases with similar substrate preference. Interestingly, the knockout cell lines seem to exhibit heightened general proteolytic response to IM stimulation compared to the wildtype cells (Fig. 2-12A–B). However, PrAMA analysis suggests that this increased activity likely arises from proteases other than ADAM10 or ADAM17. Fig. 2-12C shows that wildtype cells cleave the good ADAM substrates (6 & 9) at a greater relative rate compared to the knockout cell lines, even if the absolute cleavage rate is lower. Consequently, PrAMA infers the knockout cells to exhibit stronger MMP9 rather than ADAM activity.

Ultimately, PrAMA results that show ADAM17 to be activated in response to PMA agree with multiple other reports in the literature (Sahin et al., 2004; Horiuchi et al., 2007; Xu et al., 2010). Furthermore, PrAMA indications that IM stimulates ADAM10 activity also agree with previous literature (Horiuchi et al., 2007). This work complements these previous reports by observing specific ADAM17 and ADAM10 activity in a non-invasive, real-time manner, without resorting to pharmacological or genetic perturbations.

Optimal substrate selection — Various logistical constraints may exist that limit the number of substrates available to be used for PrAMA in certain applications. To address this issue, we implemented a common optimal design criterion for selecting substrates so as to maximize PrAMA accuracy. The determinant of the covariance error matrix $\mathbf{R}_M$ is one metric describing the volume of inference uncertainty. Minimizing this volume, which is equivalent to maximizing the determinant of its inverse and minimizing the condition number, ultimately reduces model uncertainty and optimizes PrAMA inference accuracy (Belsley et al., 1991). Using the following function, we optimally selected subsets of the total 18 FRET-substrates to perform various

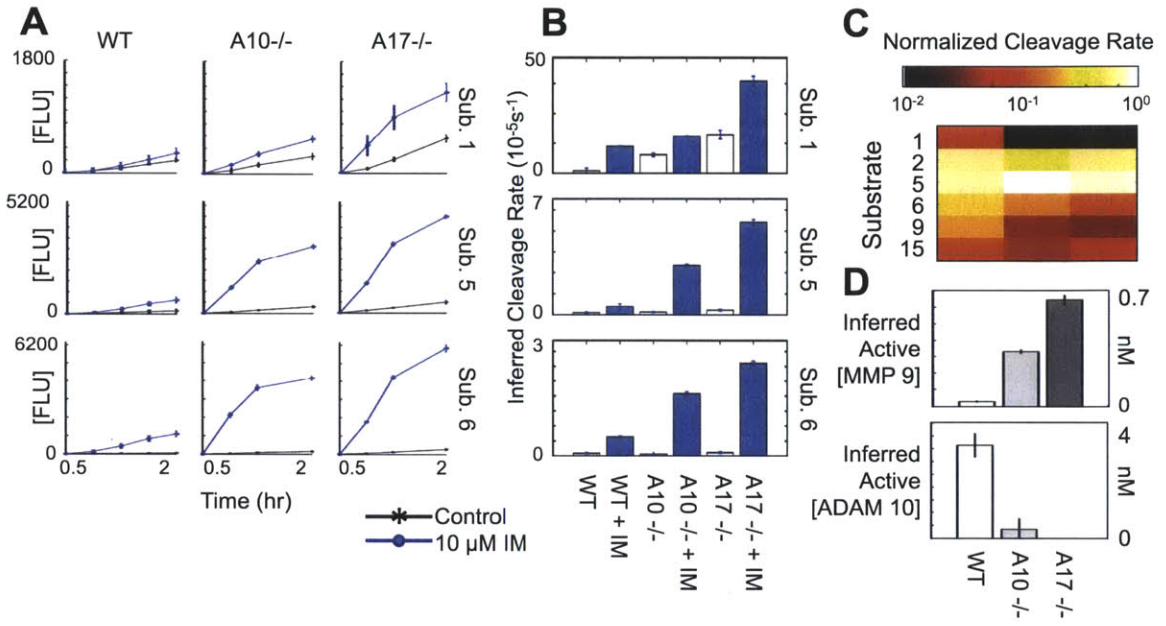

Figure 2-12: **Live cell inference of ionomycin stimulated MP activity.** PrAMA was conducted using three cell lines (WT, ADAM10-/-, and ADAM17-/- MEFs) and 6 total substrates, tracking substrate cleavage up to two hours after adding substrate. (A) Time-lapse fluorimetry for 3 of the 6 total substrates used in this experiment. (B) Inferred substrate cleavage rates ($V_0/[S]_0$), corresponding by row to the time-courses shown in A and using all time-points in A for the inference. Stars indicate $p<0.05$, comparing between cleavage rates for the control and stimulated conditions. (C) Observed cleavage vector $\mathbf{V_0}$ (columns), normalized to have total signal of 1, for each of the three cell lines. (D) PrAMA inference results for the increased activity caused by IM stimulation, using significance threshold σ_T =1.4. No significant increase in MMP2, MMP14, or ADAM17 activity was detected at this threshold. For MMP9 and ADAM10, all inferred differences were statistically significant ($p<0.05$). For all subplots, error bars indicate standard deviation of three biological replicates.

PrAMA inferences:

$$\min det(\mathbf{R_M}) = \max det(\mathbf{C}^T\mathbf{R}_D\mathbf{C}) \tag{2.9}$$

We tested the substrate selection strategy on PrAMA inference of mixtures involving MMPs 1, 2, 3, 7, and 8, where only those 5 enzymes are considered in the parameter matrix $\mathbf{C}$ (Fig. 2-13A-E). In general, the impact of optimal substrate selection increases as the number of substrates decreases. Surprisingly, optimal substrate selection allows PrAMA to use the theoretically minimum number of substrates, equal to the number of MMPs considered in the parameter matrix $\mathbf{C}$, without significantly impacting inference accuracy. PrAMA inference of double-enzyme mixtures containing ADAM10, ADAM17, MMP2, and MMP14 maintains an accuracy indicated by the area under the ROC curve (AUROC) of roughly 0.9 even as the number of substrates decreases by nearly 50% (Fig. 2-13F). In general, the optimal combination of substrates depends on which MPs are being analyzed by PrAMA, and the optimal substrates combination for distinguishing ADAM activities is distinct from the optimal combination for distinguishing activities of MMPs 1–8.

2.4 Discussion

Existing techniques used to study MPs each have advantages and disadvantages. Zymography is one of the oldest and most common MP activity assays, and a variety of zymographic techniques exist to measure the activities of diverse proteases and their inhibitors (Kleiner et al., 1994; Fowlkes et al., 1997; Oliver et al., 1997). Most zymographic techniques involve SDS-PAGE electrophoresis, which prevents continuous real-time measurement (and generally disrupts non-covalent MP complexes). *In situ* zymography, often applied to frozen tissue sections, allows for the observation of localized protease activity (Galis et al., 1995). However, in most cases the substrates used for *in situ* zymography (e.g., gelatin) are readily cleaved by a variety of MPs and measurements consequently lack specificity. Recently a variety of methods have been developed to observe the proteolytic degradation of endogenous MP substrates. For example, mass spectrometric techniques can quantify hundreds of proteins that

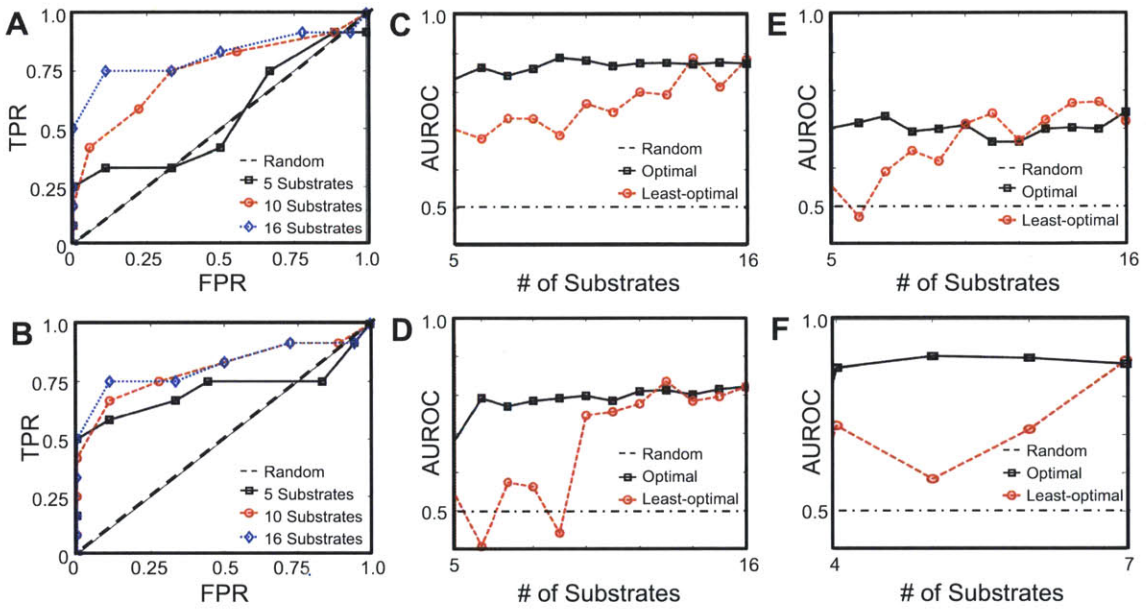

Figure 2-13: **Optimal substrates selection improves PrAMA accuracy.** (A,B) ROC curves describing PrAMA accuracy for double-enzyme experiments involving MMPs 1,2,3,7, and 8, where PrAMA uses a worst (A) or best (B) subset of the substrates, as defined in the text. (C–F) Area under the ROC curve (AUROC) as a function of the number of best (black) or worst (red) substrates, for various sets of PrAMA experiments: single (C), double (D), and triple (E) enzyme mixtures involving MMPs 1,2,3,7, and 8; (F) double-enzyme mixtures involving ADAMs 10 and 17.

have freshly cleaved amide bonds within complex biological samples (Butler et al., 2010). Problems involving the complex relationship between proteases and their substrates make it difficult to accurately and non-invasively infer the contributions of specific MPs to the global patterns of endogenous substrate degradation. Yet another recently developed method, activity based probes (ABPs), support the direct and specific measurement of diverse protease activities (Saghatelian et al., 2004; Blum et al., 2005; Sieber et al., 2006). Like typical zymography substrates, however, ABPs can act on a broad range of related proteases. Therefore, assays involving ABPs face a trade-off between invasiveness and specificity. For specific protease identities to be ascertained, biological samples analyzed with ABPs can be resolved by size through electrophoresis.

Synthetic polypeptide protease substrates have been developed for an increasingly wide range of enzymes. Within the last few years several FRET-substrates have been designed with some specificity, thereby supporting their application in complex biological samples (Alvarez et al., 2005; Moss and Rasmussen, 2007; Xu et al., 2010). Nevertheless, cross-reactivity with closely related MPs and distantly-related, but much more non-specific, proteases can still complicate the interpretation of FRET-substrate activity assays. As an example, several FRET-substrates with some specificity for ADAM17 have recently been developed with a sequence based on the ADAM17 cleavage site on pro-TNFα (Alvarez et al., 2005; Hiraoka et al., 2008; Mohammed et al., 2004; Moss and Rasmussen, 2007). Multiple recent reports employ these ADAM17 FRET-substrates, even though they have documented cross-reactivity with related MPs (Walker et al., 2009; Caescu et al., 2009). At least six MPs have been recognized to cleave endogenous pro-TNFα, in some cases at the same site (Moss and Rasmussen, 2007; Amour et al., 2002; Haro et al., 2000). Such non-specificity complicates interpretation of the observed FRET-substrate cleavage, especially when comparing multiple correlated MP activities in the same biological sample (Walker et al., 2009). Several MPs cleave many of the same synthetic and endogenous substrates that ADAM17 cleaves, suggesting the repertoire of ADAM17 substrates could be a subset of the repertoire for more promiscuous MPs (e.g., MMP14) or non-specific

proteases like plasmin (Tam et al., 2004). At least to some degree, this situation is conceivable not just for ADAM17 but for a variety of MPs, and would make identifying truly specific substrates impossible.

Although MPs have been extensively studied for decades, no method yet exists to assay multiple protease activities in real-time with high specificity and non-invasiveness. One explanation partly accounting for this fact is that the ubiquitous regulatory interactions, diverse substrates, and distinct roles played by closely related MPs have only recently become fully appreciated. Both MMPs and ADAMs engage in regulatory networks controlled by cyclical feedback interactions. For example, ADAMs participate in an autocrine positive feedback loop in mammary epithelial cells: EGFR transactivation stimulates Erk activity, which in turn stimulates ADAM shedding of EGF ligands, further activating EGFR (Joslin et al., 2007). In this situation, common methods of ascertaining the influence of ADAM activity on EGFR signaling, such as by applying a protease inhibitor or siRNA treatment, can both disrupt the underlying feedback interactions and potentially create compensatory reactions whereby closely related ADAMs modify their activity to accommodate perturbations (Sahin et al., 2004; Le et al., 2009; Blobel et al., 2005; Joslin et al., 2010). As another example, many MMPs activate themselves and one another. Such interactions can create positive feedback interactions that allow, for example, an initiating MMP activation event to trigger further protease activation (Sternlicht et al., 2001). We predict non-invasive, multiplexed, real-time, and specific measurements of MP activity will be critical towards understanding the complex regulatory mechanisms underlying MP networks.

We anticipate that PrAMA should have broad applicability in protease biology. FRET-substrates have been extensively used for high-throughput inhibitor screening with individual purified enzymes. PrAMA would allow inhibitor screening to be performed in more complex enzyme mixtures and biological samples, and could be adapted for high-throughput *in vitro* functional assays of inhibitor activity. As discussed above, PrAMA is well suited for network-level analysis of *in vitro* protease activity, and PrAMA can scale up and down in scope depending on the particular

application. At the most basic level, PrAMA could use multiple FRET-substrates in tandem to bolster the specificity of an activity measurement for even a single protease. In other words, the parameter matrix $\mathbf{C}$ could be as small as (2 substrates $\times$ 1 enzyme). PrAMA can capture protease activity on a variety of time-scales, depending on the particular application. We demonstrate high sensitivity measurements that are made over the course of >5hrs, and live-cell measurements can detect significant differences in cleavage within 30min. Dynamic measurements on this short time-scale can be relevant for detecting rapid post-translational protease activation, while longer time-scale measurements have relevance, for example, to phenotypic responses that are downstream of transcriptional changes. Soluble FRET-substrates can be directly applied to both live-cells and cell lysate for protease activity measurement (Alvarez et al., 2005; Walker et al., 2009). Our initial experiments show that PrAMA can operate by adding individual yet distinct FRET-substrates to live-cells in a multi-well format. Furthermore, FRET-substrates with distinctive excitation/emission spectra may be simultaneously combined in the same solution for PrAMA of a single biological sample. FRET-substrates have been tethered directly to 3D substrata such as collagen (Packard et al., 2009), providing localized measurement of protease activity. For simultaneously analyzing many protease activities, the mathematical framework behind PrAMA can be applied to microarrays of peptides, for instance, that contain hundreds or thousands of FRET-peptide substrates. Previous work with peptide microarrays has demonstrated how patterns of peptidase activity can be deconvoluted using non-linear least squares in a similar manner to PrAMA, ultimately to infer the presence of specific proteases within complex biological samples (Gosalia et al., 2006). PrAMA builds upon this previous work by using non-invasive dynamic measurements of peptide cleavage rather than static snapshots of cleavage patterns to quantitatively infer kinetic parameters (i.e., $\mathbf{V_0}$). The advantages of PrAMA's mathematical framework include explicitly accounting for substrate depletion & photobleaching effects, as well as readily extending to dynamic measurements of protease activity where $\mathbf{V_0}$ is not constant. Strategies for selecting optimal substrates depend on the application. Most commonly, optimal substrates are individually chosen by the combination

of their specificity and cleavage efficiency (Gosalia et al., 2006; Caescu et al., 2009; Chen et al., 2002). In this work, however, we employ a global optimization strategy to identify the set of substrates that combine to yield the greatest specificity and the least inference uncertainty. The principles behind PrAMA, including strategies for optimally selecting substrates, are readily extendable to other classes of enzymes, such as caspases and cathepsins. FRET-based protease substrates have been successfully applied to measuring *in vitro* caspase activation. Like MPs, however, individual caspases have overlapping substrate specificity and it can be difficult to interpret which specific caspase has become activated (Bouchier et al., 2008). Lastly, PrAMA inference has many potential uses involving clinical samples. For example, simultaneous measurement of multiple protease activities in patient fluid samples or biopsies could reveal mechanistic insight and/or identify activity-based markers of disease state for diagnostic/prognostic use. Ultimately, this work presents an integrated mathematical and experimental framework that can be adapted and extended to a broad range of applications. We have demonstrated various methods of *a priori* analyzing how best to design PrAMA experiments, whether it be through choosing optimal substrates, identifying which proteases can be specifically measured with the available substrates, or understanding how to account for experimental variability.

Chapter 3

Multiplexed Protease Activity Assay for Low Volume Clinical Samples Using Droplet Based Microfluidics and Its Application to Endometriosis

Abstract

As principal degrading enzymes of the extracellular matrix, metalloproteinases contribute to various pathologies and represent a family of promising drug targets and biomarker candidates. However, multiple proteases and endogenous inhibitors interact to govern metalloproteinase activity, often leading to highly context-dependent protease function that unfortunately has impeded their clinical utility. We present a method for rapidly assessing the activity of multiple specific proteases in small volumes (<20ul) of complex biological fluids such as clinical samples which are only available in very limited amounts. We have developed a droplet-based microfluidic platform that injects the sample into thousands of picoliter-scale droplets from a bar-

coded droplet library containing mixtures of unique moderately selective FRET-based protease substrates and specific inhibitors and monitors hundreds of the reactions thus initiated simultaneously by tracking these droplets. Specific protease activities in the sample are then inferred from the reaction rates using a deconvolution technique, Proteolytic Activity Matrix Analysis (PrAMA). Using a nine-member droplet library with three inhibitors and four FRET substrates, we apply the method to the peritoneal fluid of subjects with and without the invasive disease of endometriosis. Results show clear and physiologically relevant differences with disease (in particular, decreased MMP-2 and ADAM-9 activities).

3.1 Introduction

Extracellular proteases participate in myriad physiological and disease processes, most prominently by degrading extracellular matrix components. In particular, matrix metalloproteinases (MMPs) and A Disintegrin and Metalloproteinases (ADAMs) have been investigated as potential drug targets and diagnostic biomarkers. Metalloproteinase activities are regulated through a tight network of multiple proteolytic enzymes and inhibitors (especially Tissue Inhibitors of Metalloproteinases, TIMPs), frequently resulting in highly context-dependent behavior that has hampered their usefulness in the clinic. Existing approaches such as zymography (Kleiner et al., 1994), activity-based enzyme-linked immunosorbent assays (ELISAs) (Lauer-Fields et al., 2004), peptide microarrays (Gosalia et al., 2006), and activity-based probes (Sieber et al., 2006) have been limited by trade-offs including throughput, simultaneous measurement of multiple activities (multiplexing), cost, and direct kinetic measurement. Alternatively, FRET-based polypeptides have been used in recently developed techniques (Gosalia et al., 2006; Miller et al., 2011) including Proteolytic Activity Matrix Analysis (PrAMA) to simultaneously ascertain multiple specific protease activities (Miller et al., 2011). The PrAMA technique interprets reaction rates from panels of moderately selective fluorogenic substrates combined with specific protease inhibitors to infer a profile of protease activities from relatively unprocessed

physiological samples. Unfortunately, this approach involves performing separate parallel biochemical reactions and consequently carries large liquid-handling and material requirements, presenting a challenge in clinical applications with limited sample quantities.

In this work we report the development and use of an integrated droplet-based microfluidics platform for initiating and simultaneously observing hundreds of protease enzyme activity reactions for hours (up to around eight hundred individual droplets using nine different reaction conditions) using limited quantities (<20 ul) of biological/clinical samples and then deconvolving the observed reaction rates using PrAMA. Compartmentalization of chemical reagents in picoliter-scale aqueous droplets allows for a potential 10^6-fold reduction in reagent consumption compared to standard methods, and facilitates the rapid monitoring of thousands of droplets, each of which may contain unique experimental conditions (Agresti et al., 2010; Brouzes et al., 2009). Droplet-based technology has recently been applied to a variety of biological applications (Zheng et al., 2003; Miller et al., 2012c; Chen et al., 2011), and pico-injectors have recently been developed to efficiently perform multistep experiments for large-scale multiplexing (Abate et al., 2010). Integration of these capabilities with PrAMA confers particular synergy: the droplet microfluidics create large scale parallel measurements of multiple protease activity reactions, while PrAMA efficiently interprets the high-dimensional kinetic data to infer multiple specific proteolytic activities.

We applied this method to study the invasive disease of endometriosis, which is generally defined by the presence of endometrial-like tissue residing outside the uterus and strongly associates with pain and infertility. Metalloproteinases have been implicated as important enzymes in endometriosis (Szamatowicz et al., 2002), but their activities in the context of dysregulated endogenous inhibitors remain less clear (Szamatowicz et al., 2002; Gilabert-Estelles, et al., 2003). Using the droplet based multiplexed activity assay, we were able to analyze minimal amounts of clinically-obtained peritoneal fluid from patients with and without endometriosis, and found distinct patterns of protease activity between disease and control samples. In particular, we discovered that MMP-2 and ADAM-9 enzymatic activity decreased with

disease and concluded that MMP and inhibitor (TIMP) protein concentrations alone failed to accurately describe the altered proteolytic turnover of specific enzymes. The multiplexing capability achieved through the microfluidic assay not only improved discrimination between control and disease samples, but also supported inference of multiple, specific protease activities that otherwise would have been ambiguous or sample-limited using traditional approaches.

3.2 Results

The complete method developed is schematized in Fig. 3-1A. The details of microfluidic device design, fabrication and operation are described in Sec. B.1.We first prepared protease substrate libraries consisting of 50 um diameter, monodisperse water-in-oil emulsions using droplet generator chips. We formulated droplets to encapsulate unique biochemical assays comprising aqueous solutions of particular protease substrates and, in some cases, protease inhibitors. The PrAMA methodology describes strategies for optimally selecting panels of substrates and inhibitors for accurately inferring specific protease activities. In brief, multiple unique FRET-substrates with distinct enzyme selectively profiles can be utilized in parallel to permit computational inference of specific enzyme activities. This inference can be additionally strengthened by incorporating the comparison of reaction rates in the presence or absence of specific inhibitors. In this application, we identify specific droplet compositions by optically "barcoding" them using specific concentrations of one or more indicator dyes (Alexa-405 and Alexa-546) (Agresti et al., 2010; Brouzes et al., 2009). The barcoded droplets are stabilized using an oil-phase surfactant, and then are mixed in a single tube where they remain stable for more than one week. Once the droplet library has been generated, it is flowed into to the device (Fig. 3-1B) where individual, single-file droplets (containing protease substrates) are mixed 1:1 with a fixed volume of biological sample (containing proteases) using a pico-injector (Fig. 3-1C) (Abate et al., 2010). After mixing with the sample, droplets are flowed to an integrated incubation chamber (Fig. 3-1D) where they are monitored via time-lapse fluorescence

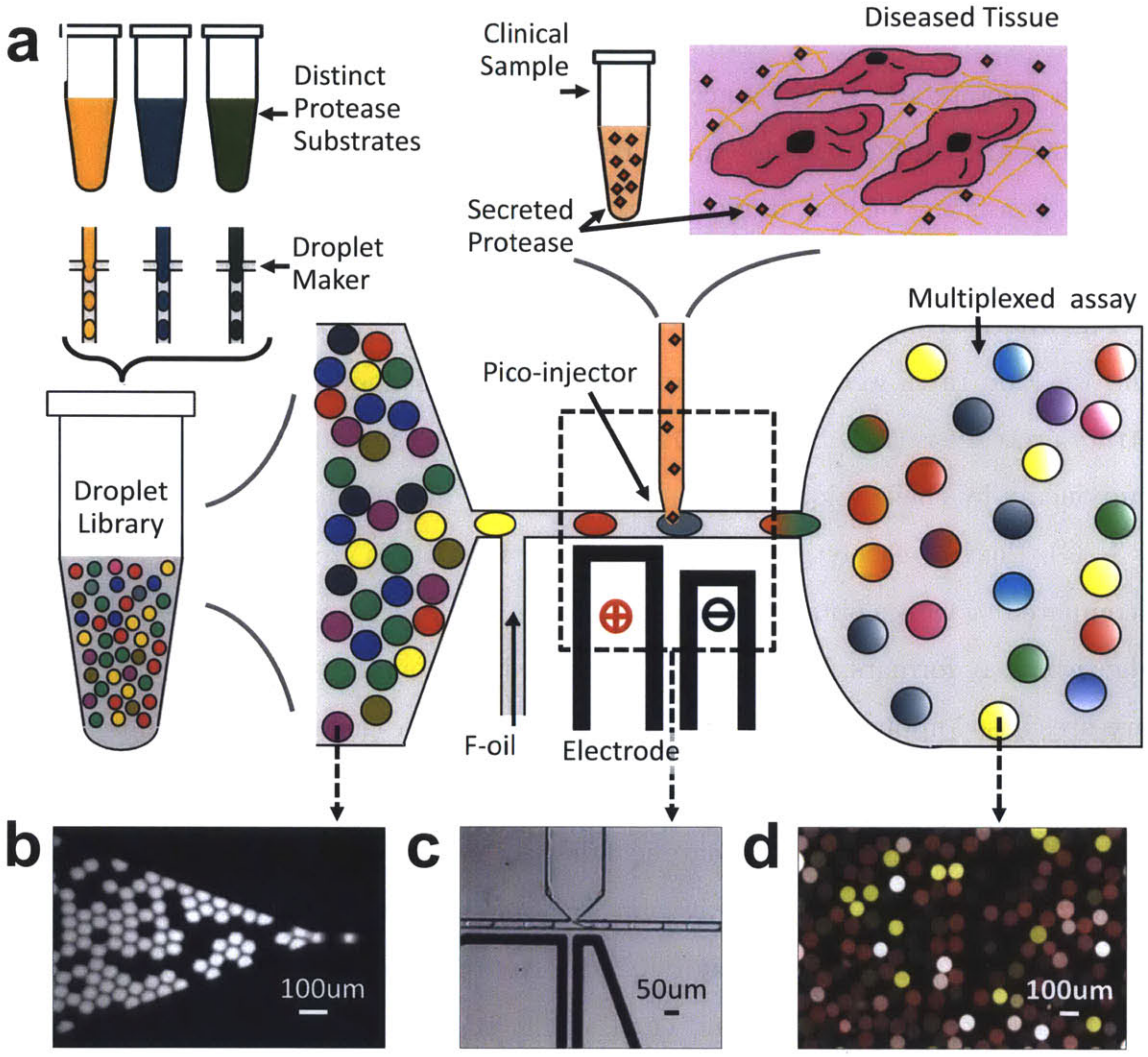

Figure 3-1: **The microfluidic device.** (a) Various protease activity substrates, and in some cases inhibitors, are encapsulated into the droplets distinguished by optical dye labeling. (b) The droplet library flows into a pico-injector device (c) with high volume fraction order, where they merge with the biological sample. (d) Using automated image processing, the protease activity reactions in each droplet are tracked over time in an observation chamber.

microscopy for over three hours. Hundreds of droplet reactions can be simultaneously monitored for hours using automated droplet tracking software, enabling both multiplexed capacity and accurate inference of reaction rates across multiple replicate droplets.

To establish microfluidic PrAMA accuracy, we first conducted reactions using recombinant enzymes. Purified enzyme solutions were injected into a four-component droplet library, consisting of four unique protease substrates (Sec. B.2) that were barcoded with Alexa-405. Sample-injected droplets were fluorescently imaged for 1.5 hours (Fig. 3-2A). Droplet tracking software (Sec. B.3) interpreted time-course images, and reaction rates were inferred from the increase in fluorescence resulting from substrate proteolysis (Fig. 3-2B) (Miller et al., 2011). We gated the four droplet library components by their unique groupings of indicator fluorescence and reaction rate (Fig. 3-2C). Enzyme catalytic efficiencies inferred from these groupings (Fig. 3-2D) were compared to values obtained using a standard plate-reader assay (R^2 >95% between the two assay formats, Fig. B-1A). We inferred the composition of unknown enzyme mixtures based upon their observed patterns of substrate cleavage, and results indicated >95% accuracy (Fig. 3-2E-F, Fig. B-1B-C).

After using purified enzymes to validate the device, we applied it to the in vitro study of an immortalized cell line (12Z) established from a peritoneal endometriotic biopsy11. To ascertain the proteolytic activity response of these cells to TNF-a (an implicated inflammatory cytokine; Grund et al., 2008; Banu et al., 2008), we stimulated cells for 24hrs, collected and clarified supernatant, and analyzed the samples with the aforementioned four-component library (containing substrates S1-S4, Fig. 2G). Standard methods show here and in previous work that MMP-2, MMP-9, and ADAM-9 are all secreted by 12Zs (Grund et al., 2008; Banu et al., 2008). We used microfluidic PrAMA to specifically analyze MMP-2, MMP-9, ADAM-9, and ADAM-10 (a known substrate of ADAM-9; Moss et al., 2011) activities. Of these enzymes, microfluidic PrAMA detected a significant increase only in ADAM-9 activity with TNF-a treatment (Fig. 3-2H, $p<0.01$, bootstrapping test; Miller et al., 2011). To validate this result, we conducted activity assays in the presence or absence of the

specific ADAM-9 inhibitor, proADAM-9 (Fig. 3-2I) (Moss et al., 2011). The results, combined with ADAM-9 ELISA data (Fig. 3-2J), confirmed upregulation of active ADAM-9 secretion ($p<0.05$). By comparing PrAMA to ELISA results, we found roughly 50% of ADAM-9 to be active, compared to <10% of MMP-2. This discrepancy is consistent with high observed concentrations of the inhibitor TIMP-2, which does not inhibit ADAM-9 but does inhibit MMP-2 (Fig. B-2) (Amour et al., 2002).

We then analyzed peritoneal fluid (PF) from patients with moderate/severe endometriosis (n=7) and compared them to PF from a control population without the disease (n=6) (Sec. B.4). Microfluidic PrAMA inference of specific protease activities revealed significant differences between disease and control samples. PF lines the pelvic cavity and comprises a heterogeneous mixture of leukocytes, cell debris, thousands of soluble proteins, and likely over 100 proteases and protease inhibitors that interact with endometriotic lesions (Amon et al., 2010). We analyzed clarified PF samples using a nine-component substrate library consisting of the same four substrates used previously, but with the inclusion of droplets containing a broad spectrum metalloproteinase inhibitor (BB94), pro-domain inhibitors for ADAM-9 and ADAM-10 (Moss et al., 2007; Moss et al., 2011), and buffer controls for the pro-domain inhibitors. The nine types of droplets were distinguished by a ratio of two indicator dyes (Fig. 3-3A). Overall, the observed reaction rates (Fig. 3-3B) showed strongest activity with substrate S4, which can be efficiently cleaved by both MMPs and ADAMs (Fig. B-3A,C). Addition of the broad-spectrum inhibitor BB94 reduced the observed reaction rates by 90% on average, thereby confirming S4 cleavage to be principally the result of metalloproteinases. For most PF samples, droplets containing pro-domain inhibitors for ADAM-9 and ADAM-10 exhibited significantly lower reaction rates compared to their buffer controls. Across all samples, the ADAM-9 and ADAM-10 inhibitors reduced reaction rates by an average of 25% (σ=16%) and 80% (σ=3%), respectively (Fig. B-3D). For six of the thirteen PF samples, the summed proADAM-9/proADAM-10 inhibition accounted for roughly all observed S4 cleavage. We conducted PrAMA to infer specific protease activities from the cleavage measurements (Sec. B.3), and found ADAM-10 to be the most active protease in general. Of

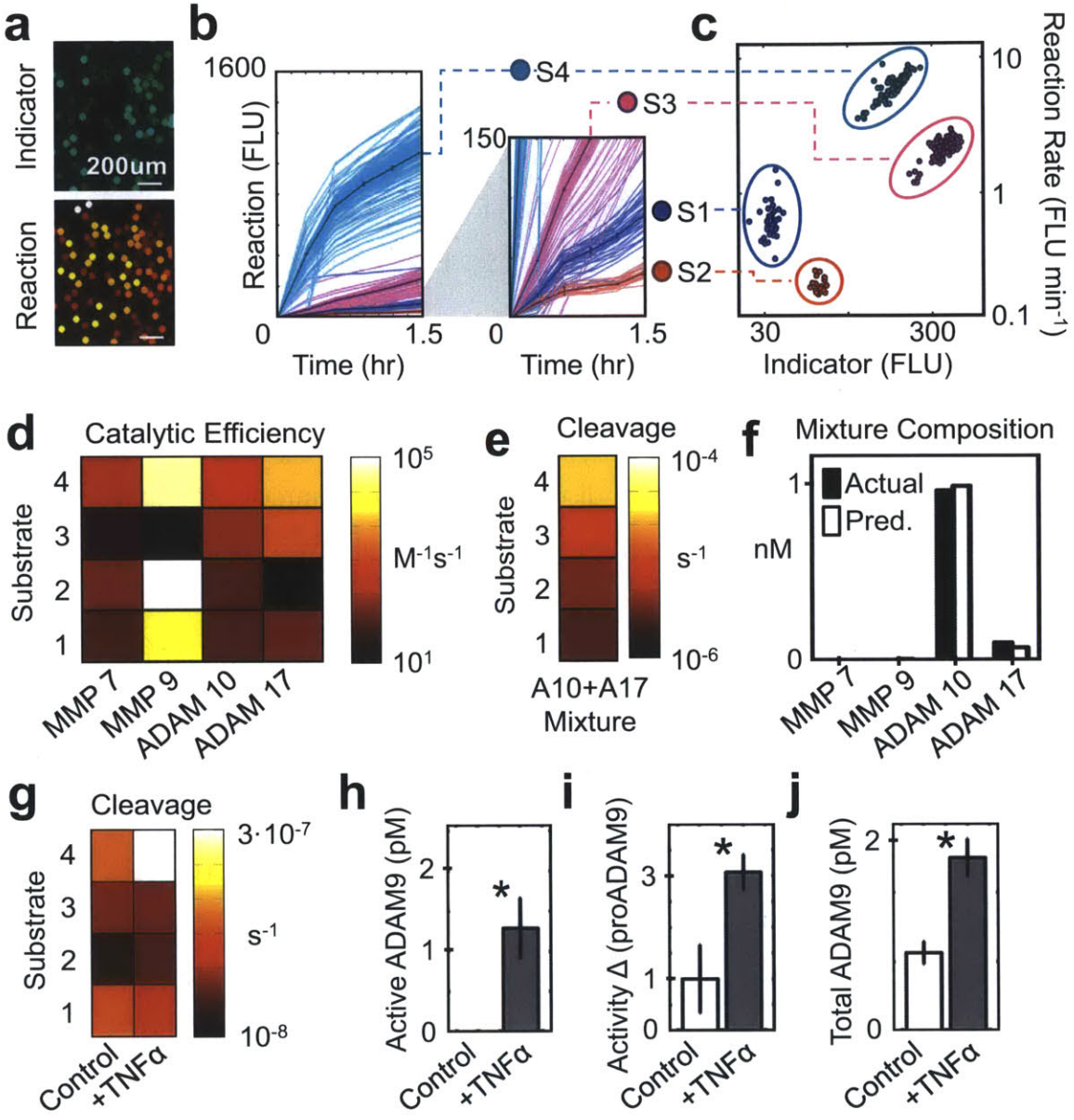

Figure 3-2: **Application to purified enzymes and cell-based assays.** (a) Droplet image of indicator dye and reaction product fluorescence. (b) Proteolysis of substrates S1-S4 was monitored by fluorescence increase. Thick lines and error bars denote mean +/- S.E.M, after individual droplets (thin lines) were classified (c) by their indicator dye and reaction rate. (d) Catalytic efficiencies were determined for samples with known concentrations of purified recombinant enzyme. (e) Observed cleavage rates for a mixture of recombinant ADAM-10 and ADAM-17. (f) PrAMA inferred the enzyme mixture from (e), using parameters from (d). (g) Substrate cleavage measured from 12Z conditioned media. (h) ADAM-9 PrAMA results corresponding to (g). (i) The difference in substrate S4 cleavage measured in the presence or absence of proADAM-9. (j) ELISA result corresponding to g-i (*$p<0.05$).

the MMPs, PrAMA results suggested MMP-2 activity to be the highest on average (Fig. B-3E-G).

Significant correlation among the observed droplet reaction rates compelled multivariate statistical approaches for data interpretation. We used partial least squares-discriminant analysis (PLS-DA) to describe patient status as a statistical function of multiple input variables, which we initially defined as the substrate cleavage measurements. Droplet data was first normalized by dividing rates by their column averages (Fig. 3-3B). In an automated manner, PLS-DA iteratively selected variables that most accurately predicted disease status, which in this case were from three reactions: S1, S4 with proADAM-9, and S4 with the buffer control for proADAM-9 (Fig. B-4A). PLS-DA combined information from these three measurements to classify disease status with 95% accuracy (Fig. 3-3C, p=0.03, permutation test).

'Normalized S1' rates statistically decreased with disease even when analyzed individually (Fig. 3-3D, p=0.042, Mann-Whitney test, Bonferroni correction). Based on the PLS-DA modeling, we also investigated the difference in reaction rates with and without the proADAM-9 inhibitor (ΔproADAM-9), and found a decrease with disease (Fig. 3-3D, p=0.035, Mann-Whitney test). Given pro-ADAM-9 specifically inhibits ADAM-9, reduced ΔproADAM-9 can be readily interpreted as a decrease in ADAM-9 activity with disease. However, interpretation of decreased S1 cleavage is less straight-forward because S1 can be cleaved by multiple proteases (particularly MMPs, including MMP-2, rather than ADAMs).

To address this issue, we utilized PrAMA inference to reveal significant differences in specific protease activities, rather than ambiguous differences in substrate cleavage, between disease and control samples. We used PLS-DA to identify the most significant PrAMA descriptors of disease status, which yielded an equally accurate (Fig. B-4B-C, 93% accuracy, p=0.04, permutation test) four-component model that ranked MMP-2 activity as the most significant determinant (Fig. B-4B). Based on the PLS-DA result, we examined relative MMP-2 activity individually and found it to decrease with disease (Fig. 3-3E, p=0.02, Mann-Whitney test). We performed ELISAs on the PF samples for MMP-2 and two of its endogenous inhibitors, TIMP-2, and TIMP-4,

to determine if changes in MMP-2 activity reflected trends in concentration (Fig. B-5). Results identified high levels of both MMP-2 and TIMP-2, but neither the absolute concentrations nor the ratio of MMP-2 to TIMP-2 significantly changed with disease. Rather, TIMP-4 concentrations significantly decreased with disease, particularly when divided by the average concentration of the three analytes for each patient sample (Fig. 3-3F, p=0.024, Mann-Whitney test, Bonferroni correction). The non-intuitive and concomitant decrease of both MMP-2 and ADAM-9 activities, in spite of reduced TIMP-4 inhibitor concentration, suggests that endometriosis perturbs multiple, overlapping protease-inhibitor interactions in the peritoneal environment. This complexity highlights the challenges associated with inferring enzyme activities from concentration alone, and emphasizes the need for multiplexed, direct activity measurements.

3.3 Discussion

In summary this work creates a platform for assessing multiple specific protease activity assays with minimal liquid handling and sample-requirement by integrating several components, including a droplet generator (Agresti et al., 2010; Brouzes et al., 2009), a pico-injector (Abate et al., 2010), and an analytical inference technique (PrAMA) (Miller et al., 2011). Accomplishing the same multiplexed measurements in a 96-well microtiter plate would consume roughly 100-fold more biological sample and reagent (20 ul sample for the multiplexed microfluidic assay, compared to a 96-well format requiring 80 ul sample/well, in triplicate, with the nine reaction conditions) which would be a prohibitively high amount especially for clinical samples. It is further noteworthy that only a tiny fraction (~30nl) of the 20 ul sample volume was actually utilized to generate droplets for multiplexed sensing, due to unoptimized world-to-chip interfacing. This leaves much room for efficiency improvement, either by drastically increasing the degree of multiplexing or further decreasing the sample volume needed. Integrated chip design that combines droplet generation with pico-injection, along with advanced droplet barcoding strategies, will be critical to

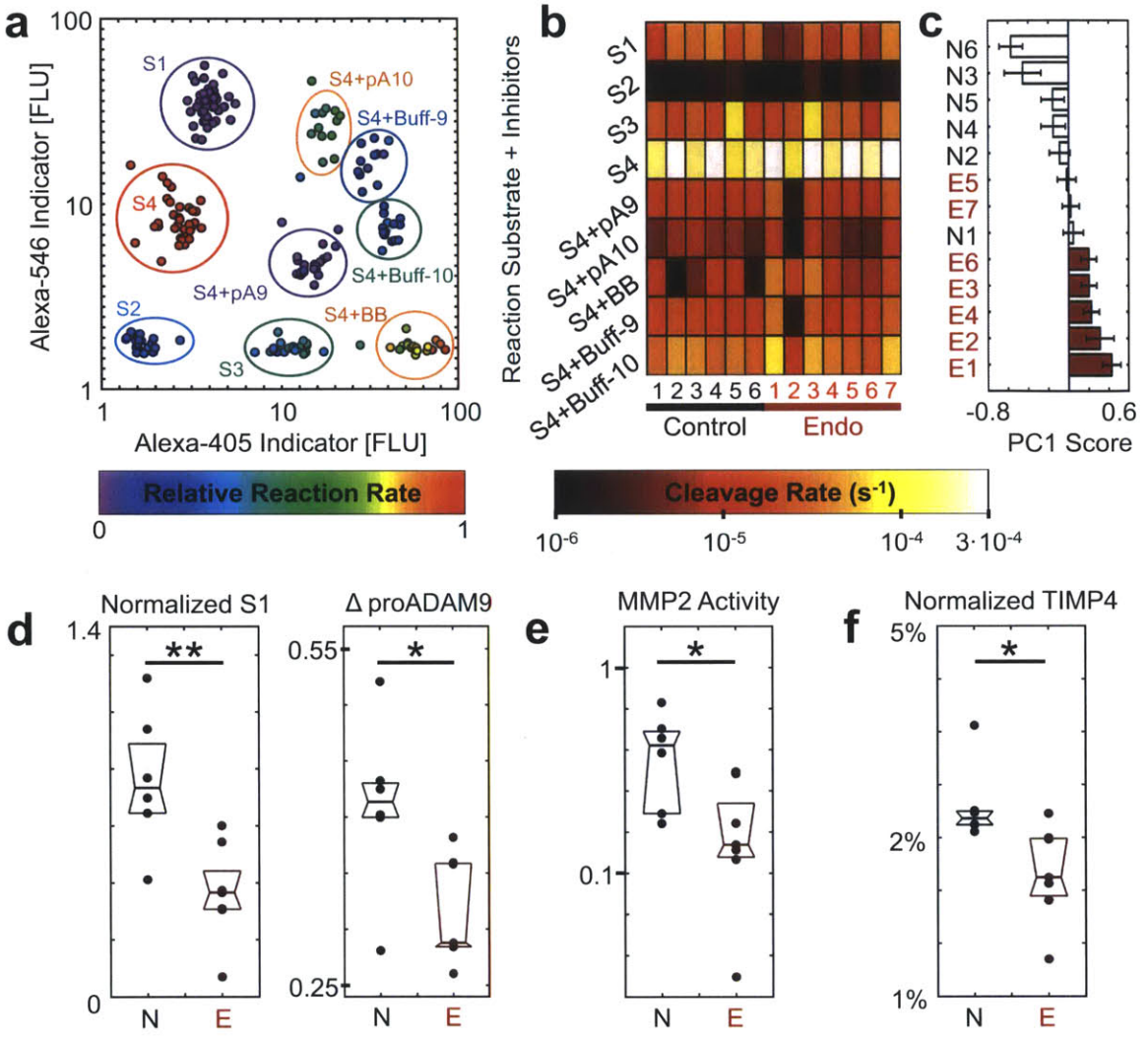

Figure 3-3: **Clinical peritoneal fluid analysis.**(a) Nine types of droplets were distinguished by two indicator dyes. (b) Peritoneal fluid samples were analyzed from six non-endometriotic patients ("Control") and seven patients with severe disease ("Endo"). The nine protease activity reactions measured for each clinical sample were defined by substrate (S1-S4) and inhibitor (pA9=proADAM-9, pA10=proADAM-10, BB=BB94, Buff = buffer controls for pro-domain inhibitors). (c) PLSDA scores plot for classifying disease status using the measurements shown in (b), +/- standard error. (d) Box-plot of the two most significant variables identified by PLSDA, which both significantly decreased in endometriotic patients. (e) The most significant PrAMA result corresponding to data in (b) determined by PLSDA. (f) TIMP-4 concentrations decreased significantly with disease, especially when normalized to MMP-2 and TIMP-2 concentrations for each patient sample ($^{**}p<0.01$,$^{*}p<0.05$).

advancing the platform multiplexing capabilities. Here we present proof-of-principle droplet libraries that utilize three barcode colors and nanoparticle dyes (Fig. B-6), and previous work has shown the potential of using microparticles, quantum dots, and hydrogels to optically distinguish potentially thousands of unique particles that have been shown as amenable to droplet encapsulation (Pregibon et al., 2007; Abate et al., 2009). While such large droplet libraries may require additional optical setup, the future challenge for enhancing the multiplexing capabilities of our platform and its overall technological potential centers on the physical generation of the droplet-library itself (Guo et al., 2012).

Our results underscore the value of multiplexing with microfluidic platforms for clinical sample analysis. When examined individually, substrate cleavage cannot generally be understood as relating to specific proteases. However, in the context of multiple reactions using inhibitors and distinct substrates, results emerged that clearly distinguished disease from control samples, while also enabling determination of specific enzyme activities. Furthermore, results from the multiplexed assay provided a novel perspective into MMP/TIMP regulation in endometriosis (Sec. B.5). Previous studies have reported conflicting observations regarding MMP-2 levels in endometriosis patients, and direct evidence of protease activity in the context of multiple interacting TIMPs has proved inconsistent (Szamatowicz et al., 2002; Gilabert-Estelles et al., 2003). Here we report that despite absence of detectable changes in MMP-2 concentration, microfluidic PrAMA revealed a significant decrease in MMP-2 activity with disease, and these observations help explain the frequent clinical observation of isolated endometriotic cysts that do not invade the surrounding tissue (Gilabert-Estelles et al., 2003). The microfluidic platform developed here could be extended to various applications (Sec. B.6), and device modularity ultimately makes the platform highly customizable for a variety of applications.

Chapter 4

ADAM-10 and -17 Regulate Cell Migration via Jnk/p38 and Opposing Feedback through Dual Ligand and Receptor Shedding in Invasive Disease

Abstract

A Disintegrin and Metalloproteinases (ADAMs) are the principal enzymes for shedding ligand and receptor tyrosine kinase (RTK) ectodomains from the cell surface. Multiple layers of activity regulation, feedback, and catalytic promiscuity impede our understanding of context-dependent ADAM function and our ability to predictably target that function in disease. This study uses combined measurement and computational modeling to examine how various growth factor environments influence ADAM activity and cellular migration in the invasive diseases of endometriosis and breast cancer. We find that ADAM-10 dynamically integrates numerous signaling pathways to direct cellular motility. Data-driven modeling reveals that induced cellular migration

is a precisely quantitative function of positive feedback through EGF-ligand release and negative feedback through RTK shedding. Although ADAM inhibition prevents autocrine ligand shedding and resultant EGFR transactivation, it also leads to an accumulation of phosphorylated HER2, HER4, and MET on the cell surface, which subsequently enhances Jnk/p38 signaling. Jnk/p38 inhibition reduces cell migration by blocking ADAM activity while additionally preventing the compensatory signaling from accumulated RTKs. In contrast, Mek inhibition reduces ADAM activity but fails to inhibit compensatory signaling from accumulated RTKs, which actually enhances cell motility in some contexts. Thus, here we present an ADAM-based mechanism of rapidly acquired resistance to Mek inhibition through reduced RTK shedding that can be overcome with combination Mek/MET inhibitor treatment. We investigate the clinical relevance of these findings using targeted proteomics of peritoneal fluid from endometriosis patients and find growth-factor driven ADAM-10 activity and MET shedding are jointly dysregulated with disease.

4.1 Introduction

A Disintegrin and Metalloproteinases (ADAMs), especially ADAM-10 and ADAM-17, are the principal mediators of proteolytic ectodomain shedding on the cell surface (Blobel et al., 2005). ADAMs and the closely related matrix metalloproteinases (MMPs) work together as "sheddases" to cleave hundreds of diverse transmembrane substrates including growth factor ligands, receptor tyrosine kinases (RTKs), adhesion molecules, and even proteases themselves from the cell surface. Unfortunately, little is known regarding how such a broad palette of proteolytic activity integrates to control behaviors such as cellular motility. Furthermore, extensive cross-talk and complexity among signaling networks, proteases, and their substrates make understanding sheddase regulation on a component-by-component basis challenging (Morrison et al., 2009). Therapeutics have targeted metalloproteinases and their substrates for the treatment of invasive diseases such as cancer, yet many metalloproteinase inhibitors have failed in clinical trials (Fingleton et al., 2008). Therefore, a need exists for

understanding how the balance of ADAM-mediated degradation integrates multiple layers of signaling networks to coordinately influence cell behavior in various disease contexts.

Here we study how sheddase activity contributes to cell migration relevant to invasive diseases such as endometriosis and breast cancer. Tumor metastasis generally involves dysregulated cell migration, with a prominent example being 'triple negative breast cancer' (TNBC) in which expression of estrogen, progesterone, and HER2/neu receptors are low or absent, but the epidermal growth factor receptor (EGFR) is frequently over-expressed (Bhargava et al., 2005; Cancer Genome Atlas Network, et al., 2012). Similarly to TNBC, endometriosis is also associated with dysregulated cell invasion into ectopic organ sites and EGFR over-expression (Uzan et al., 2009; Inagaki et al., 2007). Endometriosis is a disease defined by the presence of endometrial-like tissue residing outside the uterus, typically in the form of peritoneal lesions, ovarian endometriomas, and deep infiltrating lesions. Up to 40% of infertile women have the disease, which also exhibits co-morbidity with several cancers (Melin et al., 2008; Ozkan et al., 2008). For both TNBC and endometriosis, targeted kinase inhibitors represent attractive therapeutic strategies (Ngo et al., 2010; Hoeflich et al., 2009; Yoshino et al., 2006). However, the recent failures of several kinase inhibitors in TNBC clinical trials highlight the issues associated with context-dependent efficacy and acquired inhibitor resistance (Carey et al., 2012; Duncan et al., 2012). Many challenges in developing targeted therapeutics stem from network-level complexities such as compensatory feedback, and recent work has demonstrated how critical such mechanisms are to achieving therapeutic success, especially in cancer (Turke et al., 2010; Muranen et al., 2012). Encouragingly, computational models of systems-level biochemical networks have shown promise as tools to understand how multiple enzymatic reactions integrate to impact overall biological behavior, often with the goal of aiding the design of personalized or combination therapies (Lee et al., 2012; Miller-Jensen et al., 2007). Considering its crucial and complex role in disease, metalloproteinase regulation represents an ideal application of such network-level approaches. Here we combine recently developed multiplexed methods to directly interrogate met-

alloproteinase activities (Miller et al., 2011) with suspension array measurements of intracellular signaling and supernatant proteomics to examine how environmental cues interact with signaling and protease networks to affect cellular migration and response to inhibitor treatments. To integrate the multiple layers of experimental data, we use computational models that elucidate quantitative relationships among signaling pathways, proteases, their many substrates, and overall cell behavior. These models underscore how metalloproteinase biology must be understood as an integrated and quantitative function of multiple proteolytic reactions, and offer mechanisms of context-dependent behavior and signaling feedback. We find ADAM-10 to be a critical regulator of motility that is dynamically controlled through several signaling pathways, affecting cell-behavior through both positive feedback from EGF-ligand release and negative feedback from MET, HER2, and HER4 shedding. We find signaling inhibitors that generally reduce ADAM activity and subsequent RTK shedding, consequently allowing the accumulated RTKs to enhance downstream c-Jun kinase (Jnk) and p38 signaling. Thus, here we demonstrate an ADAM-based mechanism of rapidly acquired resistance to kinase inhibition through reduced RTK shedding that can be overcome with combination therapy. Targeted proteomic analysis of clinical samples from endometriosis patients indeed confirms growth-factor driven ADAM-10 activity and consequent MET shedding are dysregulated with disease. Overall, our results have wide implications for designing combination therapies and identifying context-dependent personalized therapeutic strategies for both kinase and protease inhibitors.

4.2 Results

Integrative network inference suggests p38 and Jnk are upstream regulators of motility that act principally through ADAM-10: We employed a variety of techniques to first characterize the signaling, sheddase, and motile responses of endometriotic cells to different growth factors that have been clinically associated with endometriosis progression (Tab. C.1). We used immortalized 12Z

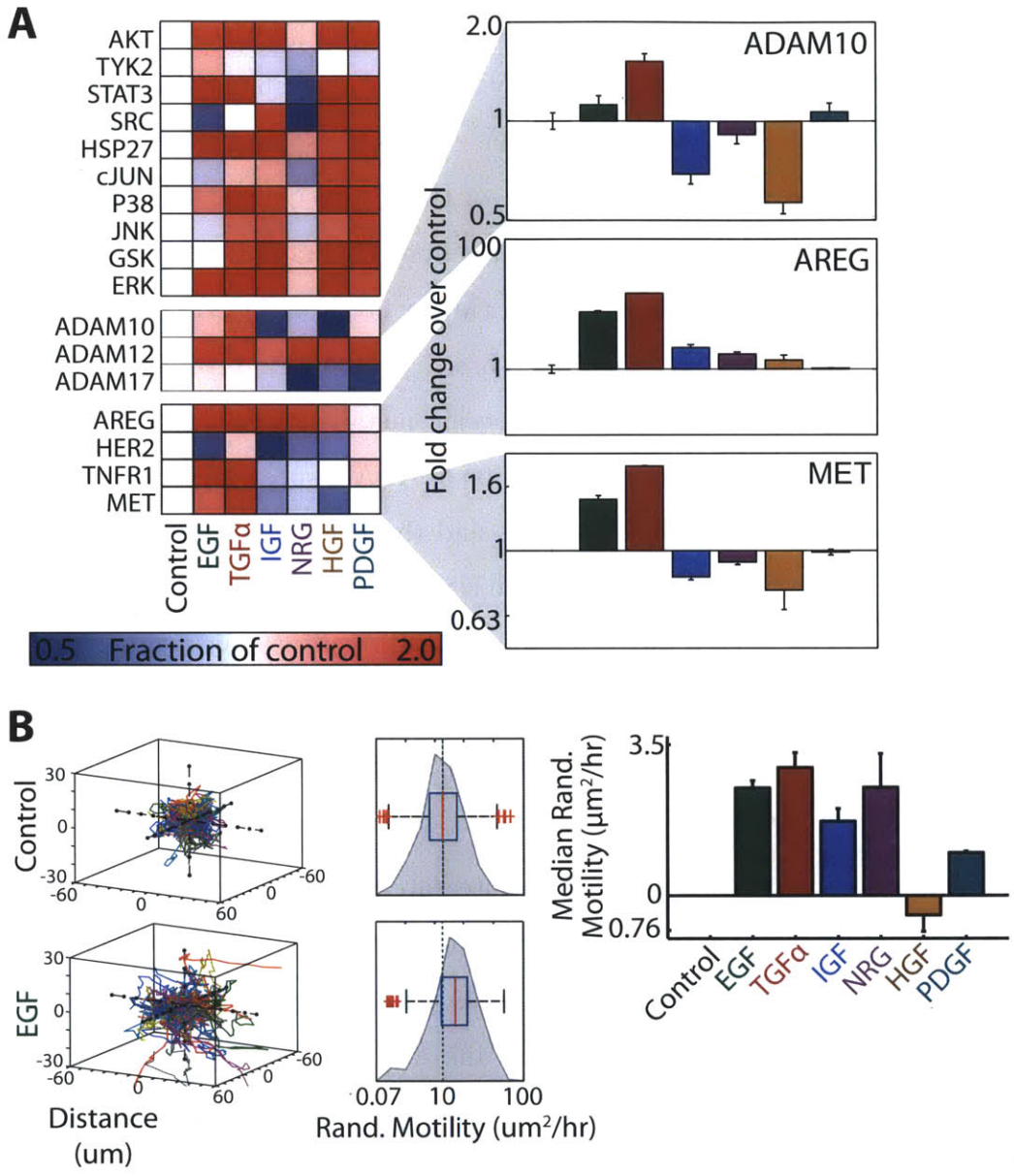

Figure 4-1: **12Z phospho-signaling, sheddase activity, and migratory response to growth factors.** (A) Normalized phosphoprotein levels, protease activities, and analyte shedding under multiple growth factor conditions. Phosphorylation was measured 5min post-stimulation. Protease activities were averaged over the first 3 hr. post-stimulation. Ligand and receptor shedding was measured 24 hr. post-stimulation. (B) Collagen-embedded cell movements were tracked via confocal microscopy over 16 hr. Individual lines represent single-cell movement tracks, after initializing positions at the origin. Box and bar plots describe median statistics.

cells, a commonly studied line established from an endometriotic biopsy (Zeitvogel et al., 2001). Shortly (5 min) after stimulating cells with six different growth factors, we quantitatively measured phosphorylation of ten key signaling proteins known to be downstream of RTK activation (Fig. 4-1A, Fig. C-1). Across the same growth factor stimulations, we recorded the proteolytic response of three principal ADAM sheddases (ADAM-10, ADAM-12, and ADAM-17) over the first 3 hours post-stimulation using Protease Activity Matrix Analysis (PrAMA) (Miller et al., 2011) (Fig. 4-1A, Fig. C-1). PrAMA uses panels of soluble FRET-based polypeptide protease substrates to directly assess the live-cell activity of specific metalloproteinases. At 24 hours post-stimulation, we quantified the accumulation of ADAM substrates that had been cleaved from the cell surface, including RTKs and the ErbB-ligand amphiregulin (AREG; Fig. 4-1A). Finally, we used time-lapse confocal microscopy to measure single-cell motile responses to growth factor stimulation. 12Z cultures suspended in collagen I gels were individually tracked for 16 hours, and the 'random motility coefficient' was calculated as a metric of single-cell motility for each condition (Meyer et al., 2012) (Fig. 4-1B, Fig. C-2 & C-3).

We used several network-inference methods to integrate these four types of data together. First, we examined correlation between all pairs of measurements as they varied across the seven treatment conditions (Fig. 4-2A). Not surprisingly, strongest correlation was found between phosphorylation of Jnk and its known substrate c-Jun. Similarly, ADAM-10 activity correlates well with accumulation of its known substrate, MET, suggesting that significant regulation of substrate shedding occurs through direct control of ADAM-10 catalytic turnover. Shedding of diverse analytes including ectodomains of transmembrane receptors and ligands correlates most positively with cellular motility, with AREG especially strongly associated (Fig. 4-2B). We performed Bayesian network analysis to integrate the data into a constrained hierarchical structure, and results suggest that Jnk, p38, and Akt signaling are the most significant upstream signaling regulators of proteolytic activity (Fig. 4-2C; Fig. C-4), acting primarily through ADAM-10. These results are further supported by complementary computational approaches, including partial least squares regression

(PLSR; Fig. C-4 & C-5). Our initial network-inference thus suggests testable hypotheses that p38 and Jnk are upstream regulators of ADAM-10 activity. ADAM-10 in turn controls multiple shedding events, most importantly AREG shedding, that ultimately direct cell motility.

Patterns of regulated sheddase activity describe 3D cell motility across multiple cell lines: Our initial network modeling found strong correlation between ligand/receptor shedding and cell motility. We tested the generality of this result by examining correlation between receptor/ligand shedding and 3D cellular motility in three EGFR-overexpressing TNBC cell lines (MDAMB231, MDAMB157, and SUM159PT). We identified significant correlation in all three cases (Fig. C-6). Because migration behavior and possibly cell signaling can substantially differ between 2D and 3D environments (Meyer et al., 2012), we compared analyte shedding in 12Z cultured on either 2D tissue culture plastic or in 3D collagen I matrices. We found significant agreement between the shedding in 2D versus 3D cultures for those species included in the Bayesian network analysis, including AREG and MET (Fig. C-7).

We used a broad-spectrum metalloproteinase inhibitor (BB94) and an EGFR blocking antibody (mab225) to determine the dependency of shed analyte accumulation on metalloproteinase activity and EGFR endocytosis of autocrine ligand (Fig. 4-3, Fig. C-8). In addition to ligands and receptors, we profiled a panel of MMPs and tissue inhibitor of metalloproteinases (TIMPs) to investigate enzymes more associated with extracellular matrix degradation. In both 12Z and MDAMB231, MMP and TIMP levels remained relatively unmoved by growth factor treatment, generally increasing only with treatment of the inflammatory cytokine TNFa in 12Z (Figs. C-8 & C-9). In contrast, we observed ligand shedding to be highly dependent on growth factor signaling and EGFR transactivation in 12Z and TNBC cell lines, suggesting a positive feedback relationship between signaling activity and EGF-ligand shedding. We performed PrAMA to examine if protease activity itself depended on EGFR stimulation and found mab225 to decrease ADAM activity (Fig. C-10 & C-11). However,

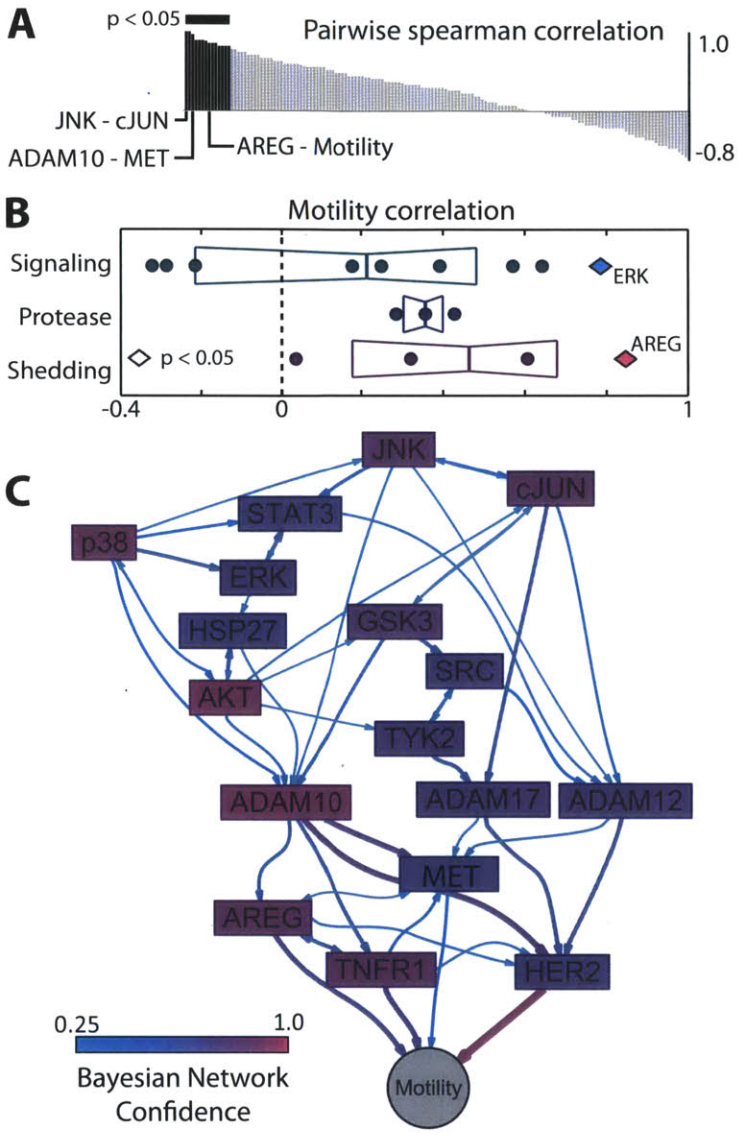

Figure 4-2: **Bayesian network analysis suggests a central role for ADAM-10.** (A) Spearman correlation among all pairs of variables in Fig. 4-1A-B. (B) Box plot showing spearman correlation between measurements in Fig. 4-1A and the observed motility in Fig. 4-1B. (C) Bayesian network analysis integrates data (Fig. 4-1) across all growth factor conditions into a hierarchical structure of directional probabilistic relationships. For *A-B*, p-values are calculated from approximation of the exact permutation distributions.

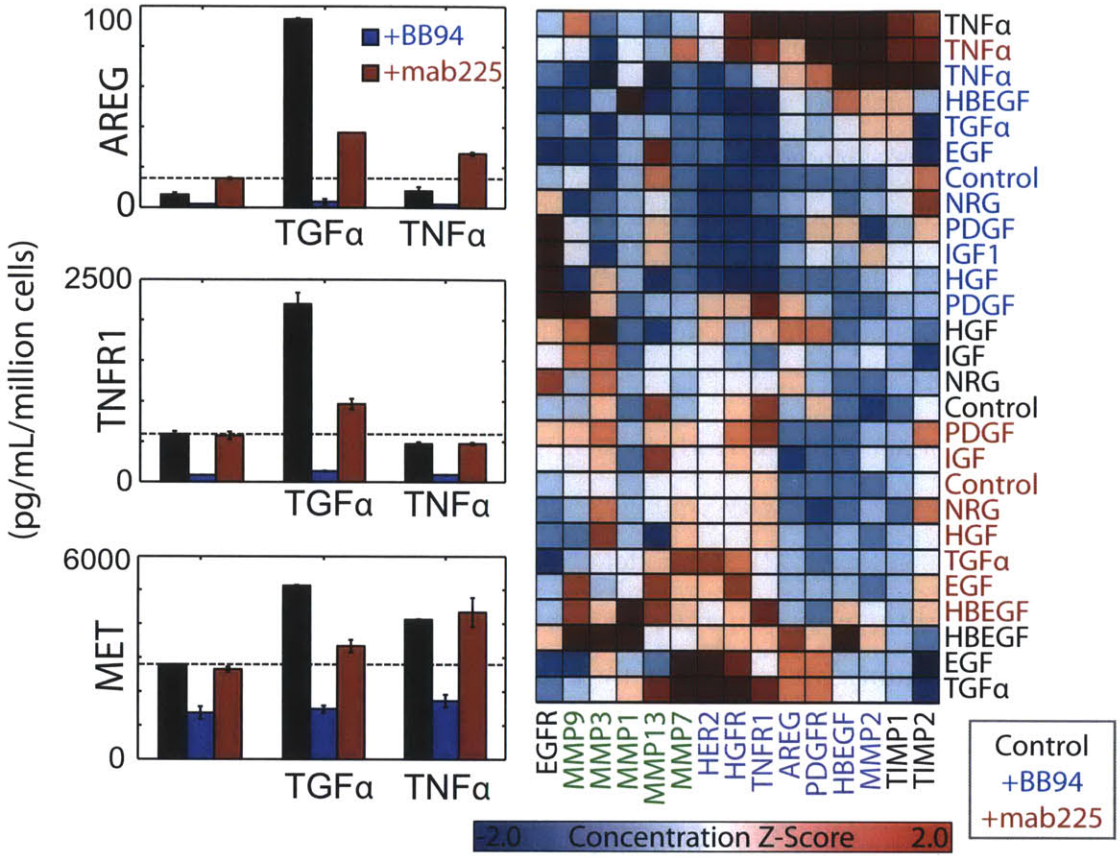

Figure 4-3: **Metalloproteinase- and autocrine EGFR- dependent supernatant accumulation.** Analyte accumulation was measured in the presence or absence of the inhibitors BB94 and mab225. Clustering reveals confirmed metalloproteinase substrates (purple) and MMPs that are upregulated by TNFa (green).

not all substrate shedding responds identically to mab225 treatment (Fig. 4-3), implicating additional mechanisms of substrate-selectivity such as trafficking and shedding by multiple proteases (Horiuchi et al., 2007).

Although individual shed analytes significantly correlate with cellular motility in multiple cell lines, single-variable relationships between shedding and motility fail to accurately predict motile responses under untested conditions in a sufficiently quantitative manner (Fig. C-12). Consequently, we implemented PLSR as a statistical method to distill the effects of multiple shedding events into key axes of control, or principal components, that quantitatively combine to describe overall cellular be-

havior. This modeling approach successfully used patterns of supernatant proteins, relying most heavily on ligand and RTK shedding, to comprehend multiple features of motility in 12Z with high accuracy (Fig. C-12). The PLSR approach also extends to TNBC cell lines, although the models generally contain cell-line specific parameters. Interestingly, 12Z and MDAMB231 cell lines exhibit sufficiently similar behavior that we could identify a unified PLSR model to accurately predict motile response to growth factors of both cell lines together with high accuracy (Fig. 4-4A). The integrated 12Z/MDAMB231 model describes cell motility as a balance of AREG, MET, and TNF receptor shedding, with the first principal component (PC-1) capturing overall ADAM proteolytic activity common among the three ADAM substrates (Fig. 4-4B). In general, however, PLSR model accuracy significantly improves with the incorporation of additional principal components (Fig. C-14), emphasizing that ligand and receptor shedding are regulated by and in turn affect multiple axes of regulation. For the integrated 12Z/MDAMB231 model, the second principal component distinguishes AREG and MET as providing competing positive and negative influences on motility, respectively. The integrated model successfully predicts that although MDAMB231 exhibit heightened motility in response to HGF, 12Z do not. AREG is the most significant model variable (Fig. 4-4C), and we find that unlike with MDAMB231, AREG does not substantially accumulate after HGF stimulation in 12Z (Fig. 4-4D). To confirm EGFR-transactivation, we found that HGF stimulated p-EGFR in MDAMB231 but not 12Z, and p-EGFR stimulation is dependent on metalloproteinase activity (Fig. 4-4E). These results demonstrate cell-type specific autocrine EGFR-transactivation, while also implicating RTK shedding as a significant modulator of distinct regulatory pathways involved in motility.

AREG-mediated EGFR transactivation sensitizes 12Z cells to kinase inhibition: With evidence of AREG-mediated EGFR transactivation, we examined whether growth factor stimulation sensitized cells to EGFR kinase inhibition. We employed an endpoint cellular invasion assay as a higher-throughput supplement that correlates with live-cell migration measurements (Fig. C-15). Briefly, cells were

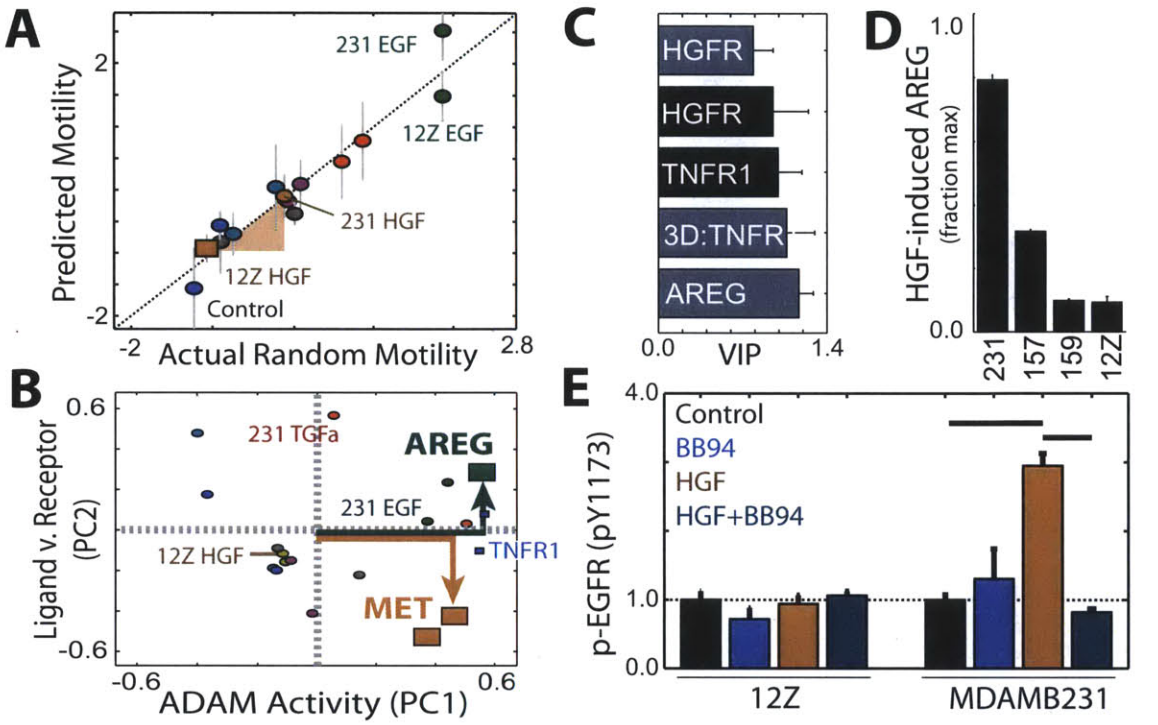

Figure 4-4: **Integrated ectodomain shedding directs motility in multiple cell lines.** (A) PSLR describes both 12Z and MDAMB231 motility as a function of analyte shedding with Q^2 fitting accuracy above 90%. The training-set of data (circles) enabled prediction of the 12Z response to HGF (square). (B) Scores (circles) and loadings (squares) for the PLSR model in *A*. (C) VIP scores for descriptor variables corresponding to *A-B*. Black bars indicate measurements consisting of the difference in receptor shedding observed with BB94 treatment. (D) HGF-induced AREG shedding is higher in MDAMB231 compared to 12Z, as a fraction of maximum AREG observed for each cell line across all conditions. (E) HGF stimulates p-EGFR in a metalloproteinase-dependent manner, in MDAMB231 but not 12Z. Bars indicate $p<0.05$, Student's t-test.

seeded under collagen I gels and exposed to bath application of growth factors after 1 hour pre-treatment with inhibitors. 24 hours after stimulation, we quantified cellular invasion into the gels (Fig. 4-5A). EGFR kinase inhibition using gefitinib and lapatinib (a dual EGFR/HER2 inhibitor) is ineffective at reducing cellular motility in unstimulated 12Z (Fig. C-16). However, nearly every tested growth factor sensitizes cells to kinase inhibition (Fig. 4-5B, Fig. C-16). We compared AREG shedding to gefitinib sensitivity across the growth factor treatment conditions and found significant correlation (Fig. 4-5B). AREG shedding is particularly enhanced with IGF1 and TNFa treatment, and IGF1 sensitizes cells most to gefitinib compared to other non-ErbB growth factors. Examination of p-EGFR in IGF1 and TNFa treated cells confirms EGFR-transactivation under these conditions (Fig. 4-5B-C). In the TNFa treated cells, EGFR transactivation depends upon autocrine EGF-ligand binding and can be blocked with mab225 treatment (Fig. 4-5C). Furthermore, TNFa stimulation enhances the effect of mab225 treatment in reducing cellular motility (Fig. 4-5D). Treatment with an anti-AREG decoy antibody effectively reduces both basal and IGF1-induced cellular motility, confirming a specific role for AREG among other potential autocrine EGF-ligands (Fig. C-16). We found the same AREG antibody also be effective in inhibiting MDAMB231 migration (Fig. C-16). In sum, this data provides further evidence for the role of AREG-mediated EGFR-transactivation in invasive cell behavior.

Ligand and RTK shedding is a function of multiple interacting proteases: Having demonstrated causal relationships between ectodomain shedding and cell motility, we next examined how shedding is regulated by proteases in these invasive disease examples. We hypothesized that ligand mediated activation may have complicated the previous network-inference results (Fig. 4-2). Therefore, we repeated experiments to measure protease activity and ligand shedding, but in the presence of mab225, across the panel of growth factor treatments. To better investigate receptor shedding, we calculated pairwise correlations both in the presence and absence of mab225. Consistent with results from the previous Bayesian and PLSR analyses,

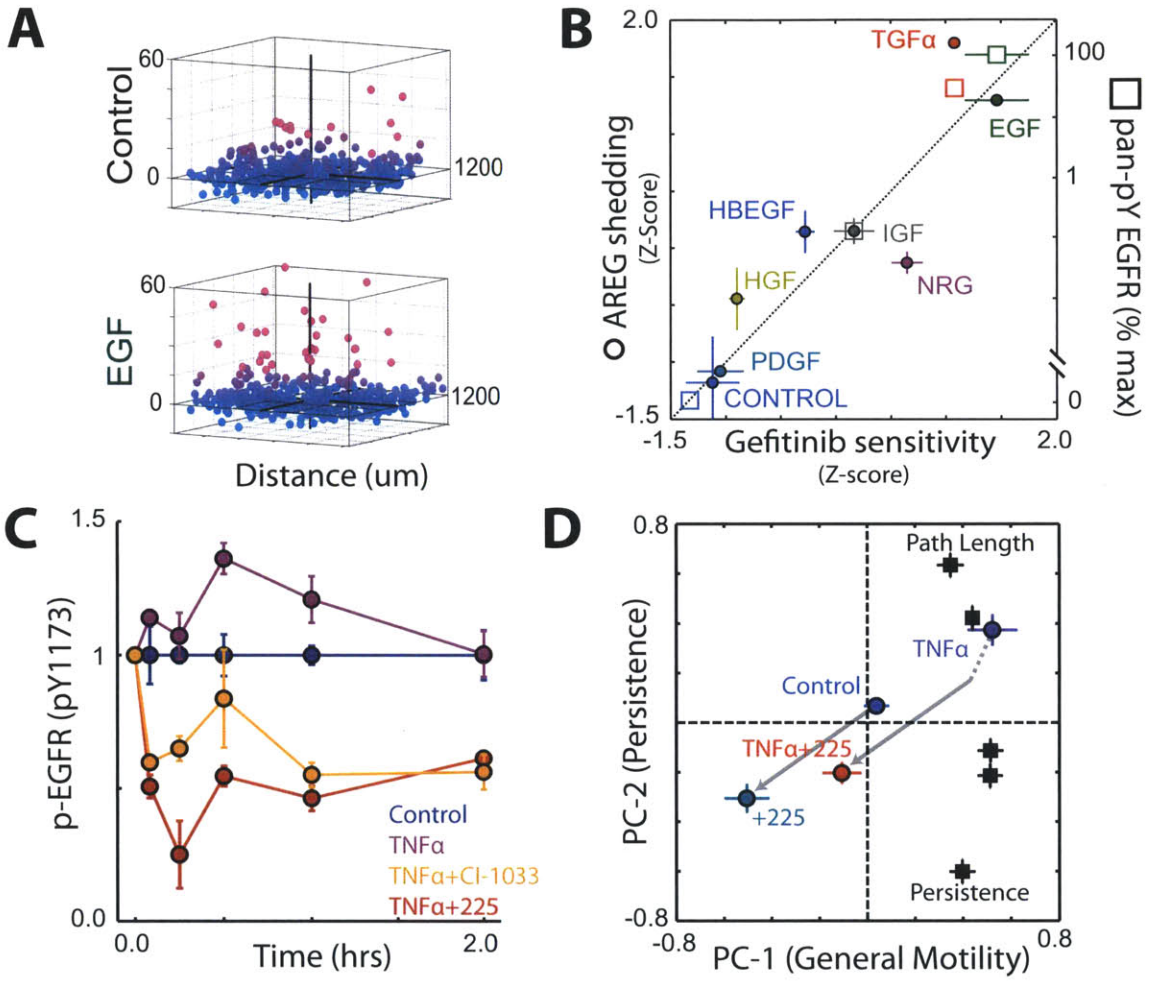

Figure 4-5: **EGFR transactivation through AREG shedding sensitizes 12Z to EGFR inhibition.** (A) 12Z cell nuclei positions reveal EGF-stimulated migration into collagen gels, as a function of distance from the plate-bottom. (B) In 12Z, growth-factor stimulated AREG shedding (left axis) and p-EGFR levels (right axis) correlate with the gefitinib efficacy in reducing cell migration. p-EGFR was not detectable in untreated 12Z. (C-D) TNFa stimulates EGFR phosphorylation (C) and motility (D) in an EGFR-ligand dependent manner. In D, cellular-motility was calculated using principal components analysis of time-lapse confocal microscopy measurements. Scores (circles) and loadings (squares) are plotted (see Fig. C-3 for details).

which did not include mab225 treatments, we identified strong correlation between ADAM-10 activity and its known substrates, HER2 and MET (Fig. 4-6A). ADAM-17 activity correlates with shedding of TNFR1, a known substrate. However, we generally found relationships between protease activities and substrate shedding to be overlapping, suggesting that cross-talk and coordinated protease regulation could play a role. We constructed PLSR models to quantitatively describe AREG shedding as a function of observed proteolytic activities (activities described in Fig. C-10 & C-11), and found that incorporating measurements of multiple protease activities, most importantly ADAM-10 and ADAM-17, greatly improves model accuracy (Fig. 4-6B, Fig. C-17). Direct inhibition of ADAM-10 and ADAM-17 proteases using siRNA treatment confirms a role for both enzymes (Fig. 4-6C, Fig. C-17). In fact, we observed that both ADAM-10 and ADAM-17 knockdown significantly affects shedding of all of the analytes measured, although ADAM-17 knockdown appears to yield a more significant reduction in substrate shedding (Fig. 4-6C, Fig. C-17). Direct examination of specific protease activities in the siRNA-treated conditions reveals ADAM-17 siRNA to also affect ADAM-10 activity (Fig. 4-6D) but not expression (Fig. C-18), which may contribute to overlapping knockdown effects on endogenous substrate shedding. Consistent with Bayesian and PLSR analyses, ADAM-10 knockdown substantially reduces the motility of both 12Z and MDAMB231 lines (Fig. 4-6E). In contrast, ADAM-17 knockdown does not show an effect in 12Z, although it does in MDAMB231, consistent with previous cell-type dependencies possibly related to non-enzymatic function (Simpson et al., 2008; Maretzky et al., 2011). Taken together, these data demonstrate significant cross-talk between ADAM-10 and ADAM-17 proteolytic activities, while also showing a substantial and consistent role for ADAM-10 in governing cellular motility.

Sheddase activity is regulated by multiple upstream signaling activities: Given that ADAM-10 plays a central role in cellular motility and is governed by multiple phosphosignaling pathways, we next interrogated the effects of direct kinase inhibition on motility. We targeted key signaling pathways identified from the Bayesian

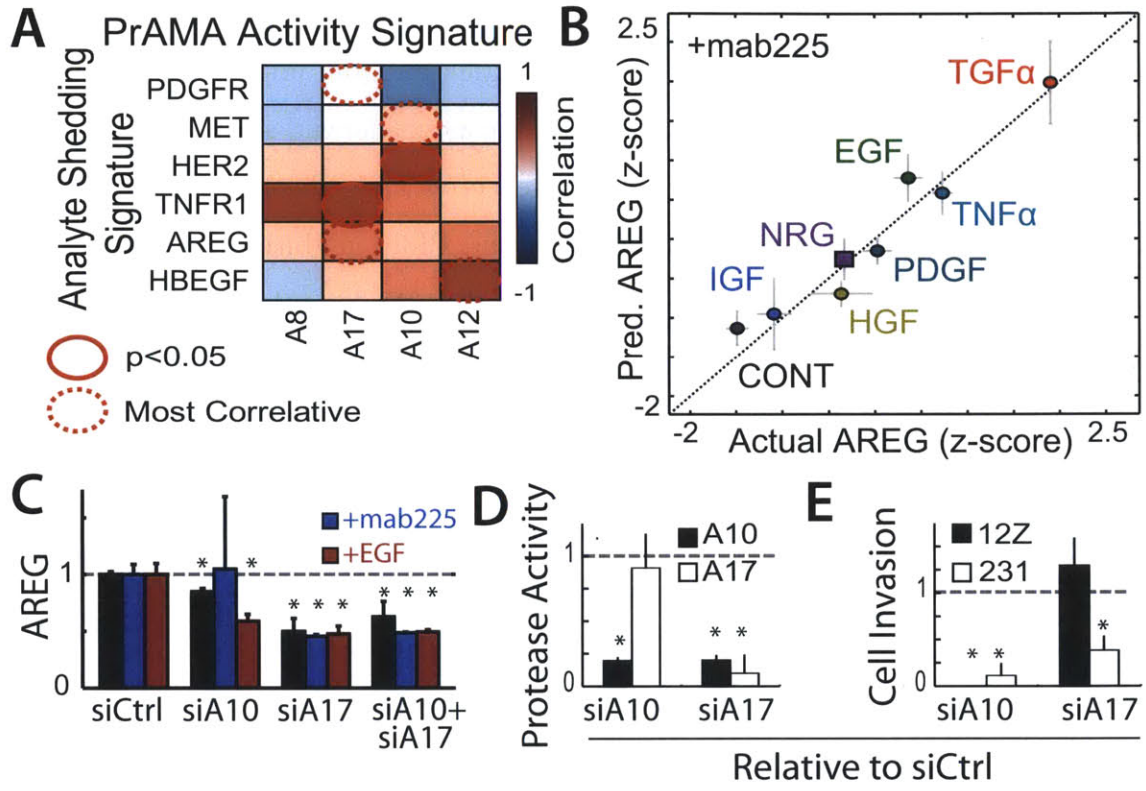

Figure 4-6: **PrAMA and siRNA reveal cross-talk among protease-substrate interactions.** (A) Pearson correlation between protease activity and shedding of analytes, including experiments with mab225 to block ligand-EGFR interaction. (B) PLSR describes AREG accumulation in the presence of mab225 as a function of ADAM activities, with Q2 accuracy > 90%. (C) Both ADAM-10 and ADAM-17 knockdown reduce EGF-stimulated AREG shedding. (D) Basal 12Z ADAM activity was measured using PrAMA. (E) ADAM-10 knockdown blocks unstimulated invasion in both 12Z and MDAMB231, while ADAM-17 only reduces invasion in MDAMB231. Asterisks indicate $p<0.05$.

network analysis and found two distinct patterns of inhibitor efficacy: p38 and Jnk inhibitors strongly reduce 12Z motility under all growth factor treatment conditions, and in contrast, Mek and PI3K inhibitors demonstrate context-dependent efficacy (Fig. 4-7A). While Mek and PI3K inhibitors effectively reduce EGF and TGFa stimulated motility, they actually enhance motile responses to NRG1b and HGF. We further tested context-dependency using more selective Mek and Jnk inhibitors and several additional cell lines. In general, the patterns of context-dependent inhibitor efficacy proved consistent: p38 and Jnk inhibition is effective under both EGF and NRG1b/HGF stimulated environments, whereas Mek and PI3K inhibition is effective only for EGFR-driven motility (Fig. 4-7B, Fig. C-19). We hypothesized that the two classes of inhibitors may differentially inhibit ADAM activity, so we directly measured protease activity in 12Z cells treated with either Mek or Jnk inhibitors. However, the two sets of inhibitors behaved similarly in blocking ADAM-10 activity (Fig. 4-7C), agreeing with the previous Bayesian network analysis. We analyzed the effect of multiple inhibitors on receptor and ligand shedding to identify potentially substrate-specific effects and found that inhibitors lack significant substrate-specificity and broadly reduce shedding of all ADAM-substrates (Fig. 4-7D).

Furthermore, we examined phosphoprotein levels of ten intracellular signaling proteins after NRG1b stimulation, in the presence of Mek or Jnk inhibitors. We implemented partial least squares discriminant analysis (PLS-DA) to identify the most significant differences in signaling network state between the two inhibitor treatments (Fig. 4-8A-B, Fig. C-20). This analysis showed that compared to Mek inhibitors, Jnk inhibitors more effectively reduce p-cJun, unsurprisingly, but also p-p38. In fact, Mek inhibition actually increases p-p38 in some cases (Fig. 4-8C).

To determine if those signaling changes arise from the concomitantly reduced ADAM activity, we measured p-p38 and p-c-Jun in the presence of BB94 and found that protease inhibition alone is sufficient to increase p-p38 and p-c-Jun (Fig. 4-9A). To explain this result, we measured phosphorylation of full-length RTK substrates of the ADAM sheddases. Protease inhibition increases p-HER2 and p-HER4 levels in 12Z (Fig. 4-9B), and we confirmed very low but significant levels of metalloproteinase-

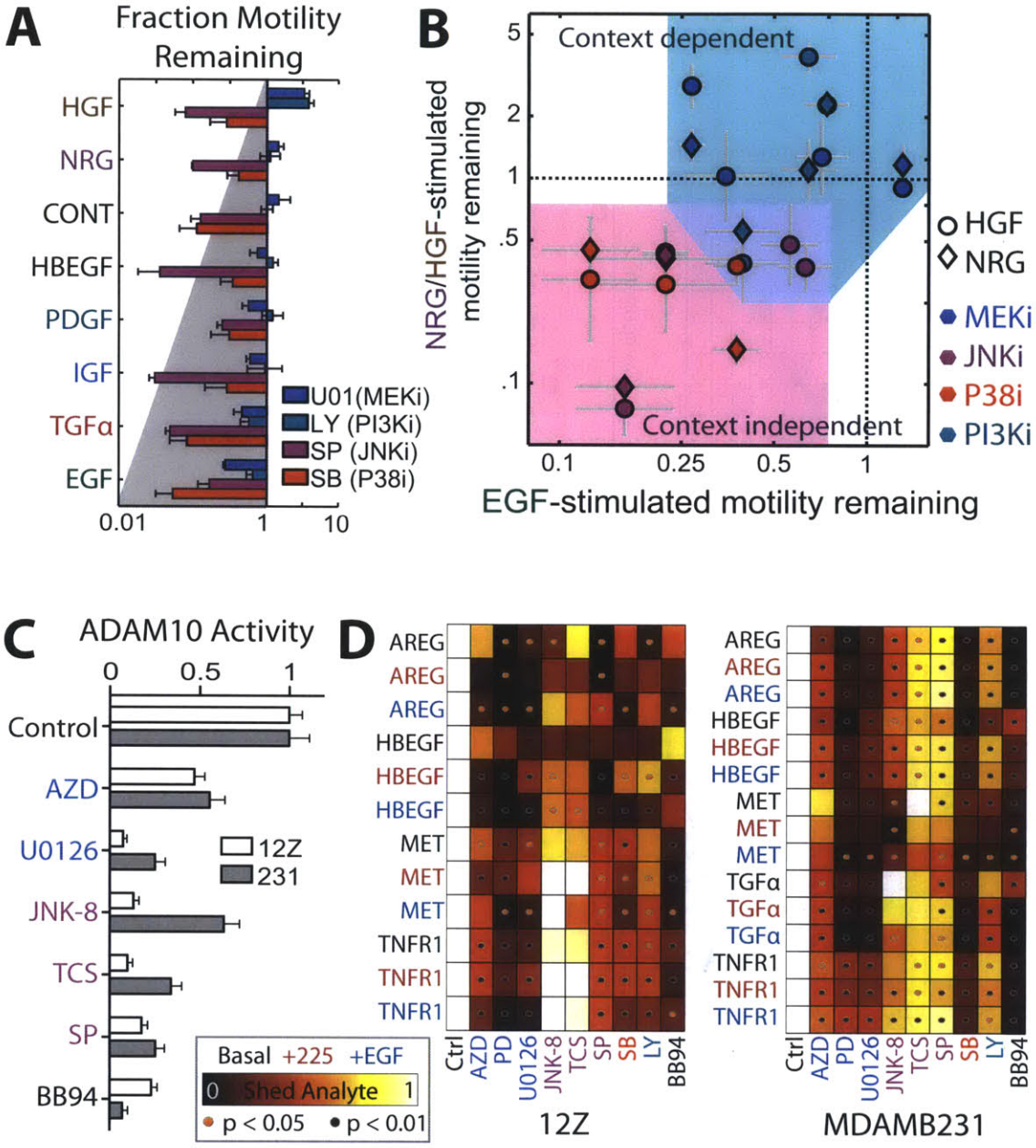

Figure 4-7: **Kinase inhibitors impact cell migration and broadly reduce ADAM activity.** (A) The effect of kinase inhibition on growth-factor induced 12Z motility. (B) Inhibitor sensitivities in multiple cell lines (12Z, MDAMB231, and SUM159PT) reveal distinct Mek/PI3K (blue) and p38/Jnk (red) context dependencies, fully labeled in Fig. C-19 . (C) PrAMA suggests Mek (blue), Jnk (purple), and metalloproteinase (BB94) inhibitors block basal ADAM-10 activity. (D) Signaling inhibitors block shedding of multiple ectodomains, measured 24 hours after inhibitor treatment in the presence of mab225 or EGF.

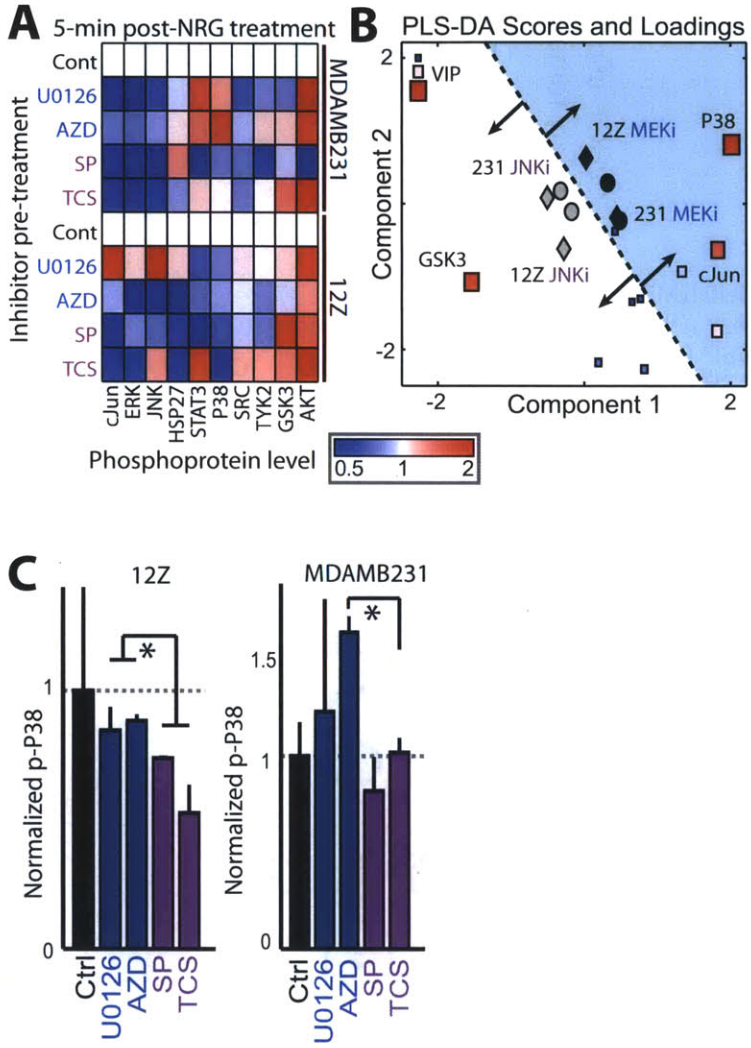

Figure 4-8: **Mek and Jnk inhibitors differentially affect p-38 signaling.** (A) Phospho-protein levels were measured 5min after NRG1b treatment, following a 1hr pre-treatment with Mek (blue) or Jnk (purple) inhibitors. (B) PLS-DA uses phospho-protein levels to discriminate Mek from Jnk inhibitor effects. PLS-DA scores (circles/diamonds) and component loading coefficients (red/blue squares) are shown, fully labeled in Fig. C-20 . (C) Jnk inhibitors (purple) more effectively reduce p-p38 compared to Mek inhibitors (blue). Asterisks denote $p<0.05$.

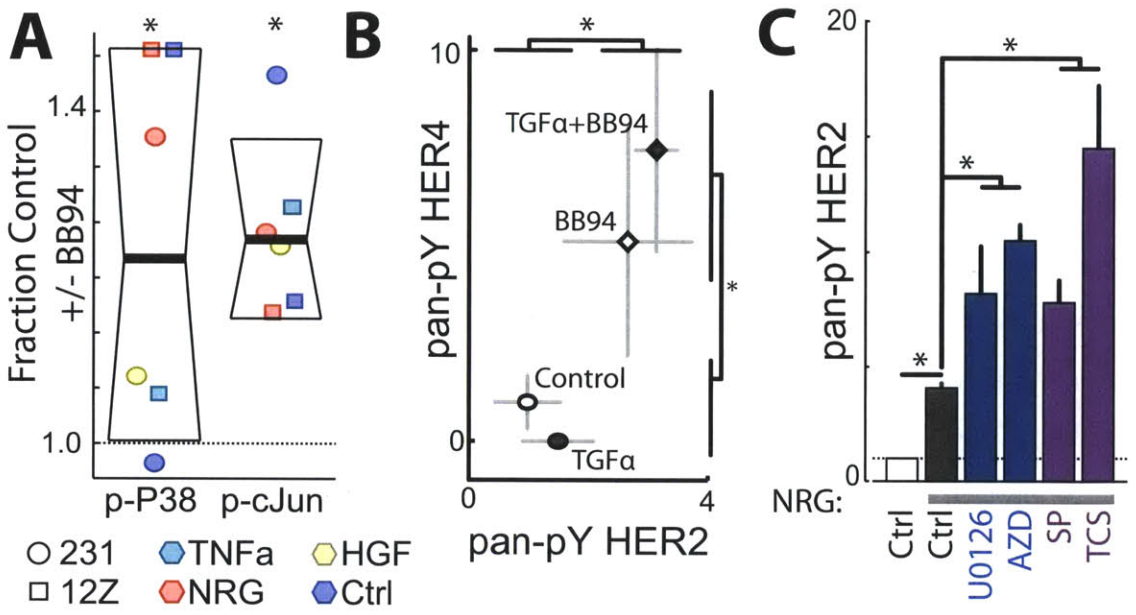

Figure 4-9: **ADAM inhibition disrupts negative signaling feedback from RTK shedding.** (A) BB94 treatment leads to increased p-cJun and p-p38 in multiple cell lines. Each data-point represents n=2 experiments. (B) BB94 treatment leads to increased p-HER2 and p-HER4 in 12Z. (C) NRG1b-induced p-HER2 increases with Mek (blue) and Jnk (purple) inhibition. In subplots *A-C*, signaling was measured 5min after growth factor stimulation, and 1hr after inhibitor pre-treatment. Asterisks denote $p<0.05$.

dependent HER4 shedding in 12Z (Fig. C-21). Finally, we investigated the effect of Mek and Jnk inhibition on NRG1b-stimulated p-HER2, and found inhibitor treatment causes increased p-HER2 levels in response to NRG1b treatment (Fig. 4-9C). We additionally confirmed that BB94 and U0126 also lead to increased p-HER2 and p-HER4 levels in MDAMB231 (Fig. C-21). Overall, these data show that the proteolytic shedding of multiple RTKs contributes to downstream Jnk and p38 signaling, which adds to EGF-ligand shedding in governing cell motility and context-dependent inhibitor sensitivity.

Clinical peritoneal fluid samples suggest dysregulated ErbB-signaling and amplified ADAM-10 activity with endometriosis: Finally, to test for relevance of our in vitro findings to in vivo pathophysiology in human patients, we analyzed surgically obtained peritoneal fluid (PF) from patients with and without endometriosis.

PF comprises a heterogeneous mixture of leukocytes, cell debris, and soluble proteins that interact with endometriotic lesions. We analyzed clarified PF samples using a targeted proteomics approach that utilized roughly the same reagents employed in 12Z supernatant profiling experiments. We assessed total protein concentrations using multiplexed and traditional enzyme-linked immunosorbent assays (ELISAs) and compared these measurements to previously reported proteolytic ADAM and MMP activities from the same patient samples (Chen et al., 2012) (Fig. 4-10A). Due to the large number of highly correlated variables measured in each patient sample, we decomposed the protein level and activity measurements into an interpretable set of principal components using principal components analysis (PCA). This unsupervised statistical method describes patient samples in terms of latent variables that capture the combined values and covariation of individual protein measurements. The first and third principal components best capture differences between control and disease PF samples (Fig. 4-10B; Fig. C-22). Interestingly, disease samples fall into two distinct clusters in principal-component space, with one cluster defined by relatively high levels of ADAM-10 activity and high concentrations of ADAM-10 substrates including AREG, HER2, and HER4. In contrast, the second cluster of disease samples exhibits relatively low ADAM-10 activity, higher levels of ADAM inhibitors (TIMPs), and higher levels of ADAM-9 activity, which is not inhibited by TIMPs. The control samples form a non-overlapping cluster between the two disease clusters. Although the sample size is small, PCA results suggest multiple disease states in endometriosis that are defined principally by dysregulation of ADAM-10 activity and corresponding changes in ADAM-10 substrate accumulation. We then used supervised PLS-DA to classify patient samples as falling into one of the three patient clusters based on a minimal number of protein measurements. PLS-DA shows that combined measurement of ADAM-9 activity and three ADAM-10 substrates (HER2, AREG, and HB-EGF) can sufficiently classify patients with high (>95%) accuracy (Fig. 4-11A). We also used PLS-DA to classify patient samples into just two groups, disease and control, and observed that combined measurements of MMP-2 activity along with MET and TIMP-1 levels sufficiently classify samples as either disease or control with

high (>95%) accuracy (Fig. 4-11B). We analyzed the simple ratio of MET to TIMP-1 levels for a more interpretable result, and observed a significant increase with disease (Fig. 4-11C). To identify the likely cellular source of increased MET shedding in the peritoneal fluid samples, we analyzed various populations of cells from healthy and endometriotic patients, including eutopic endometrial fibroblasts and peritoneal fluid mononuclear cells (PFMCs). Endometrial fibroblasts, which generally express significant MET (Yoshida et al., 2004), shed similar levels of MET compared to the 12Z (Fig. C-24). However, the PFMCs shed only 10% of MET by comparison (Fig. C-25). We used PLS-DA to identify key differences in sheddase behavior between the endometriotic 12Z and three other TNBC cell lines, and found that 12Z are most distinguished in showing substantial EGF- and TGFa-induced MET shedding (Fig. 4-11D; Fig. C-26). Experiments with primary endometrial fibroblasts demonstrate that EGF stimulates dual EGF-ligand and MET shedding in other relevant endometrial cell populations (Fig. C-24 & C-25). Therefore, both endometriotic legions and endometrial fibroblasts represent significant sources of total MET observed in the peritoneal fluid, particularly in those patients with elevated ErbB ligand present. Consequently, the ratio of MET to TIMP1 may be a good surrogate marker of ADAM-10 activity on endometrial and endometriotic tissue. Overall, these results suggest that joint dysregulation of ADAM-10 activity, ErbB-signaling, and corresponding RTK shedding play an important role in disease progression.

Combined MET-Mek inhibition blocks motility across multiple growth factor contexts: In light of the clinical evidence for enhanced MET shedding in endometriosis patients, we investigated the cellular consequences of MET shedding in 12Z. Similar to HER2 and HER4, we found that direct inhibition of metalloproteinase activity caused enhanced accumulation of both total and phosphorylated MET following HGF and NRG1b treatment (Fig. 4-12A; Fig. C-27). Given our evidence that protease inhibition can enhance MET signaling and that Mek inhibitor resistance in part arises from reduced ADAM activity, we hypothesized that Mek insensitivity in the presence of HGF and NRG1b is mediated by enhanced MET signaling. Using

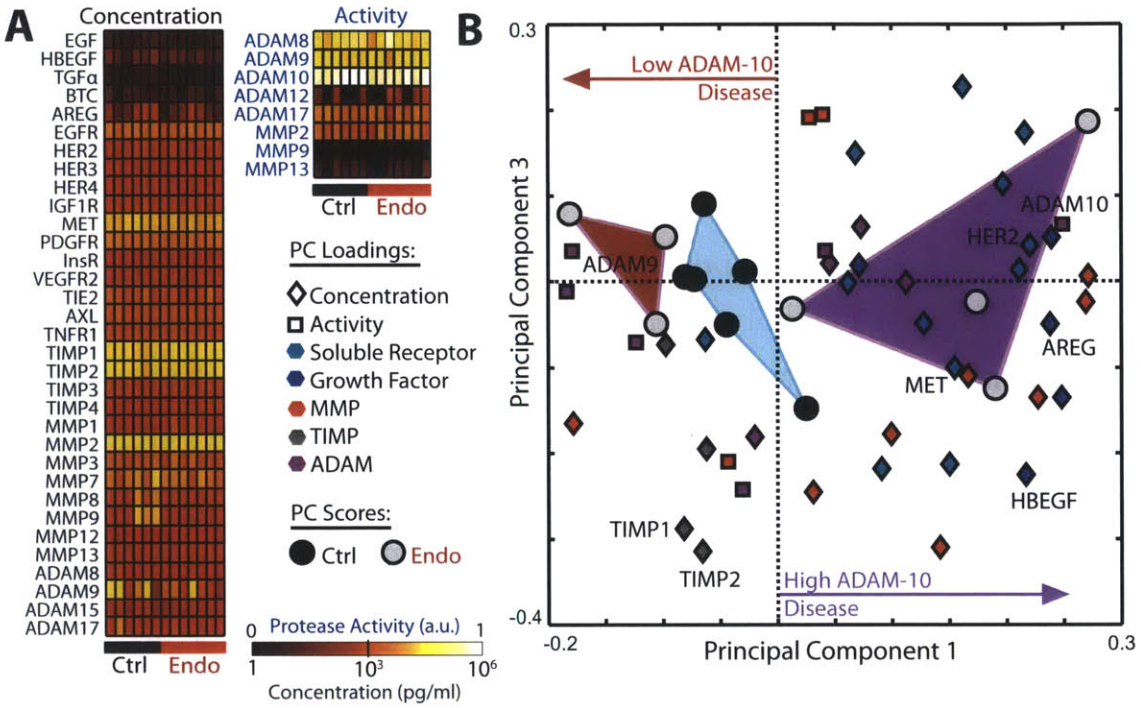

Figure 4-10: **Clinical PF samples reveal dysregulated ErbB-driven ADAM-10 activity.** (A) PF samples from patients with and without endometriosis were profiled for protein levels and protease activities. (B) PCA scores and loading plot decomposes patient samples along principal components of covariation, and separates samples into three non-overlapping clusters (fully labeled in Fig. C-22).

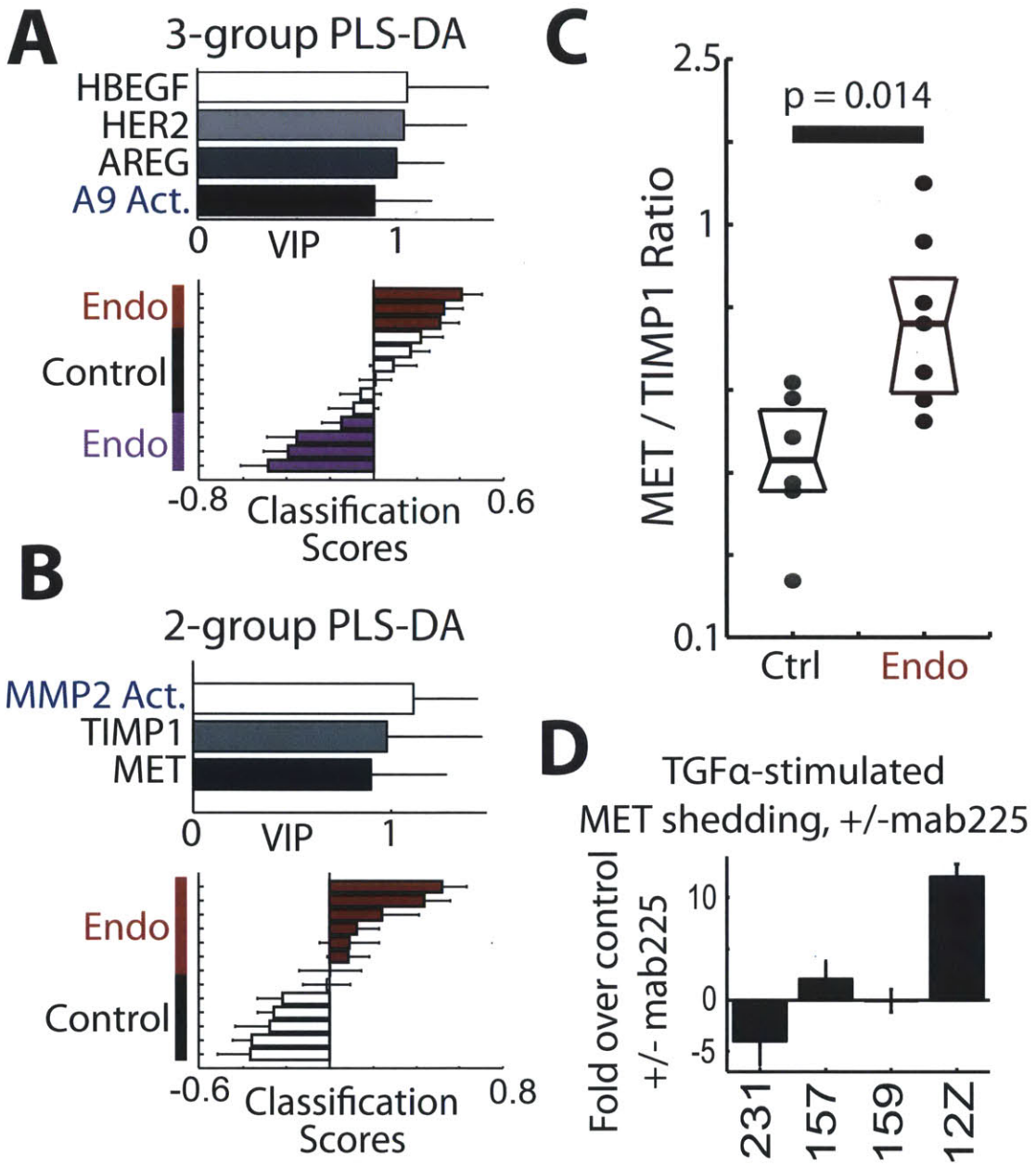

Figure 4-11: **Discriminating features within clinical PF samples.** (A) VIP scores for the four protein measurements (top) that sufficiently classify patient PF samples into the three clusters shown in Fig. 4-10B with >95% accuracy (bottom). (B) VIP scores for the three protein measurements (top) that distinguish disease from control samples with >95% accuracy (bottom). (C) The ratio of MET to TIMP1 increases with disease. (D) TGFa-stimulated MET shedding is especially pronounced in 12Z compared to several TNBC cell lines.

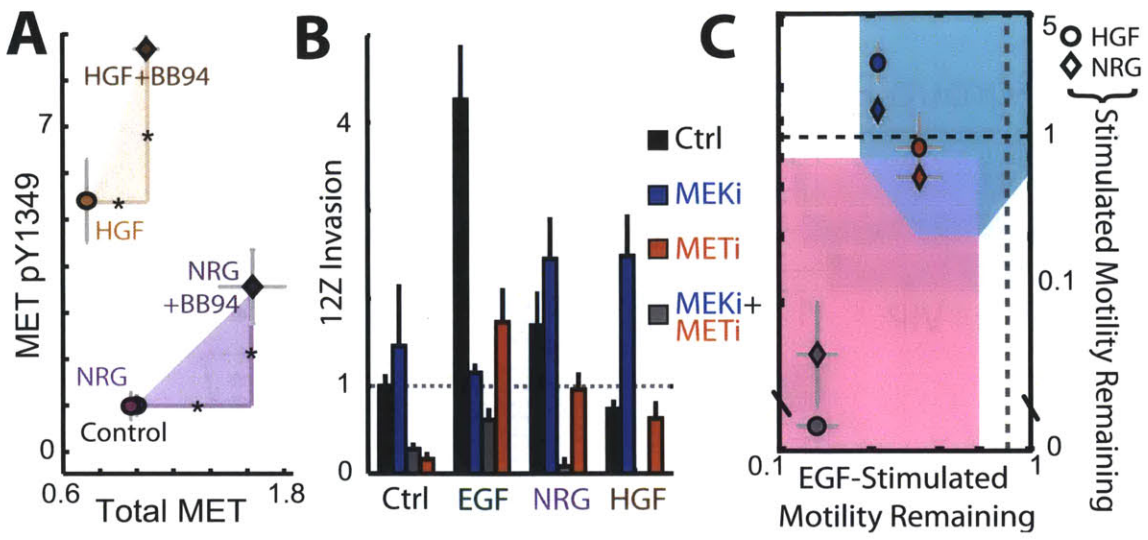

Figure 4-12: **p-MET mediates Mek-resistance via reduced ectodomain shedding.** (A) Western blot densitometry shows that BB94 increases total, full-length MET (x-axis) and pY-1349 (y-axis) in response to HGF and NRG1b. Representative gel images from replicate membranes are shown in Fig. C-27. (B-C) 1hr pretreatments with Mek and MET inhibitors reduce 12Z migratory response to growth factor stimulation in the collagen I invasion assay. (C) Combination MET/Mek inhibitor treatment exhibits greater efficacy than either treatment alone. Plot corresponds with Fig. 4-7B.

foretinib as an inhibitor of MET (and several other ADAM substrate RTKs, including VEGFR-2), we found that combination Mek/MET inhibition was more effective than either inhibitor alone, under multiple growth-factor contexts in 12Z (Fig. 4-12B-C). Overall, these results confirm the importance of alternative RTK signaling in the context of Mek inhibition and reduced RTK shedding.

4.3 Discussion

Understanding systematic regulation of ectodomain shedding has been challenging to accomplish on a component-by-component basis for multiple reasons. The web of protease-substrate interactions involves significant overlap and cross-talk: proteases (a) degrade potentially hundreds of often shared substrates (Butler et al., 2010); (b) interact with and regulate each other through direct proteolysis (Moss et al., 2011);

and (c) respond to and modulate signaling pathways (Xu et al., 2010). These multiple layers of complexity compel quantitative and multivariate approaches, and here we employ integrative experimental/computational methodologies to understand how multiple ADAM proteases interact with signaling networks to direct overall cellular behavior.

In this work we combine quantitative experimental measurements with network-inference methods to build computational models of signaling-mediated protease regulation and motility. This approach successfully identifies canonical biochemical interactions, for example between Jnk and c-Jun (Fig. 4-2A), while simultaneously providing unique insight into mechanisms of sheddase regulation. Shedding is generally considered a function of both protease catalytic activity and substrate availability, yet the balance of protease and substrate regulation remains unclear (Dang et al., 2011). Here we directly assess proteolytic turnover in a relatively "ligand-independent" manner using the PrAMA assay and find that substrate shedding generally does not precisely correlate with individual protease activity measurements, but rather represents the combined action of multiple proteases (Fig. 4-6A). Unmeasured substrate-specific effects such as endocytic trafficking likely impact overall shedding (Horiuchi et al., 2007). Nevertheless, statistical models accurately predict AREG shedding, for example, as a function of multiple protease activities including ADAM-10 and ADAM-17, without requiring measurement of AREG trafficking (Fig. 4-6B). siRNA knockdown of ADAM-10 and ADAM-17 confirm the dual dependency of AREG and indeed multiple other substrates on both ADAM-10 and ADAM-17 activities (Fig. 4-6C-D), in agreement with previous work (Le Gall et al., 2009). These results demonstrate how protease activities dynamically interact with multiple signaling pathways to govern overlapping ectodomain shedding events, and emphasize the difficulty in selectively manipulating the proteolysis of specific substrates through kinase and protease inhibitors.

Network-inference results suggest a significant role for AREG in promoting cellular motility (Fig. 4-2B), which has been relatively unexplored in endometriosis. Previous work has shown that the stability of localized intracellular signaling by sustained au-

tocrine activity promotes persistent motility (Maheshwari et al., 2001). Accordingly in this work patterns of ligand shedding correlate most closely with cellular persistence compared to other metrics of general motility (Fig. C-12). In endometriotic cells we demonstrate that AREG shedding mediates EGFR transactivation from multiple stimuli, which plays a quantitative role in determining the migratory response of cells. Comparison of EGF-ligand shedding in the presence of either mab225 or saturating exogenous EGF reveals that EGFR autocrine activity is sustained in a positive feedback loop through the EGF-induced activity of multiple ADAMs (Fig. C-11), and we identify multiple cell types including endometrial fibroblasts and PFMCs that may contribute to this feedback through EGF-ligand release (Fig. C-24 & C-25). ErbB activity significantly impacts cellular sensitivity to ErbB kinase inhibitors, and in this work we find cytokine and growth factor stimulation sensitizes endometriotic cells to gefitinib treatment through induced AREG shedding (Fig. 4-5B). In breast cancer, pre-clinical studies have shown that ADAM-10 and ADAM-17 inhibitors reduce EGF-ligand shedding and synergize with EGFR-kinase inhibitors to reduce tumor growth in xenograft models (Liu et al., 2006; Witters et al., 2008).

Although ADAM involvement in ErbB-ligand shedding makes them compelling drug targets in ErbB-driven disease, the biological consequences of ADAM-mediated RTK shedding continue to be poorly understood. In HER2+ breast cancer, ADAM-10 inhibition reduces HER2 shedding, which generally has been described as beneficially limiting the accumulation of the membrane-bound HER2 fragment (p95HER2) that remains after ectodomain proteolysis (Xia et al., 2004). However, it remains unclear how p95HER2 activity compares to full-length HER2, especially after ligand stimulation, and soluble HER2 ectodomain can inhibit signaling (Ghedini et al., 2010). For other RTKs including HER4 and MET, shedding likely reduces RTK signaling at the cell surface (Naresh et al., 2006; Foveau et al., 2009). TIMP-1 inhibition of MET shedding in breast cancer enhances MET signaling and increases liver metastasis (Schelter et al., 2011). In this work we demonstrate that cellular motility is an integrative process that depends not just on AREG shedding, but also on the combined and quantitative effect of multiple proteolytic reactions, including RTK

shedding. We find that ADAM-mediated receptor shedding downregulates HER2, HER4, and MET signaling. Reduced ADAM activity and RTK shedding, either through metalloproteinase inhibition or indirectly through signaling pathway inhibition, leads to accumulation of intact RTKs on the cell surface. RTK accumulation potentiates the signaling response to HGF and NRG1b, and causes enhanced RTK phosphorylation and downstream activation of Jnk and p38. Consequently, Mek and PI3K inhibitors actually enhance the motile response of endometriotic and breast cancer cells to NRG1b and HGF treatment by inhibiting RTK shedding while failing to block the compensatory p38 and Jnk activity that results from signaling of the un-shed RTKs. Previous studies implicate Jnk and p38 in endometriosis and cancer (Yoshino et al., 2004; Wagner et al,. 2009), and our results show that Jnk and p38 inhibitors effectively reduce ADAM-activity while also blocking the compensatory signaling and motility regardless of the growth factor environment. Overall, these results have significant implications for the design of combination therapies involving the numerous signaling pathways that affect ADAM activity, and complement previous studies that stress the importance of Jnk/p38 pathways in cell migration (Huang et al., 2003).

The emergence of secondary resistance to targeted kinase inhibition represents a major obstacle in developing successful therapeutics, and in this work we identify a novel ADAM-mediated mechanism of rapidly acquired inhibitor resistance that has potential applications for a variety of kinase and protease inhibitor therapies. Although Mek and PI3K inhibitors show efficacy in TNBC xenograft mouse models (Hoeflich et al., 2009), secondary resistance to Mek inhibitors has been well documented and arises from upregulation of RTKs that are known ADAM substrates, including PDGFRb, MET, and Axl (Duncan et al., 2012). Furthermore, the presence of growth factors that activate known ADAM-substrate RTKs, for example MET, facilitates the emergence of resistant populations (Turke et al., 2010). Consistent with these results, here we present that Mek inhibitor resistance arises through multiple upregulated RTKs, many of which have been implicated in other reports including MET and HER2. In this work we demonstrate that proteases play a role in the

acute upregulation of receptor levels, and this is particularly relevant in the presence of growth factors that have been previously implicated as chemoprotective and pro-metastatic microenvironmental cues (Turke et al., 2010; Wilson et al., 2012; Goswami et al., 2005). In endometriosis, kinase inhibitors are in the earlier stages of testing and acquired inhibitor resistance is not yet a clear problem. Nonetheless, we demonstrate that the same logic of combination therapies can be equally successful in our in vitro model for overcoming compensatory signaling pathways that arise secondarily from inhibitor treatment.

Analysis of clinical samples from endometriosis patients helped demonstrate the relevance and inherent overlap of ADAM-mediated proteolysis and RTK signaling dysregulation in disease progression. Although many previous studies have examined ErbB-signaling and metalloproteinase levels individually (Sec. C.1), here we present a multivariate analysis of systemic interaction between ErbB-ligands, RTK shedding, and metalloproteinase dysregulation. Furthermore, we use measurements from a recently developed microfluidic device to analyze protease activity directly and relate these observations to corresponding protease substrate levels observed in the same patient sample (Chen et al., 2012). Clinical results confirm many of the observations made in vitro, for example demonstrating significant correlation between ADAM-10 activity and accumulation of known ADAM-10 substrates such as HER2 and AREG (Fig. C-23). Interestingly, we find that disease samples comprise two distinct clusters defined in large part by the balance between ADAM-9 and ADAM-10 activities (Fig. 4-10B). ADAM-10 is a known ADAM-9 substrate, and ADAM-9 has been observed to downregulate ADAM-10 activity on the cell surface (Moss et al., 2011). Common among both clusters of disease samples, however, was the observation that the ratio of MET shedding to TIMP-1 concentration increased with endometriosis, thereby confirming the relevance of MET signaling in designing therapeutic strategies that may impact ADAM activity (Fig. 4-11C). Previous work has shown TIMP-1 to inhibit the establishment of endometriosis in a mouse model, and these effects were primarily assumed to be MMP-related (Bruner et al., 1997). However, in this work we demonstrate the critical role of ADAM-10 in mediating in vitro cellular invasion,

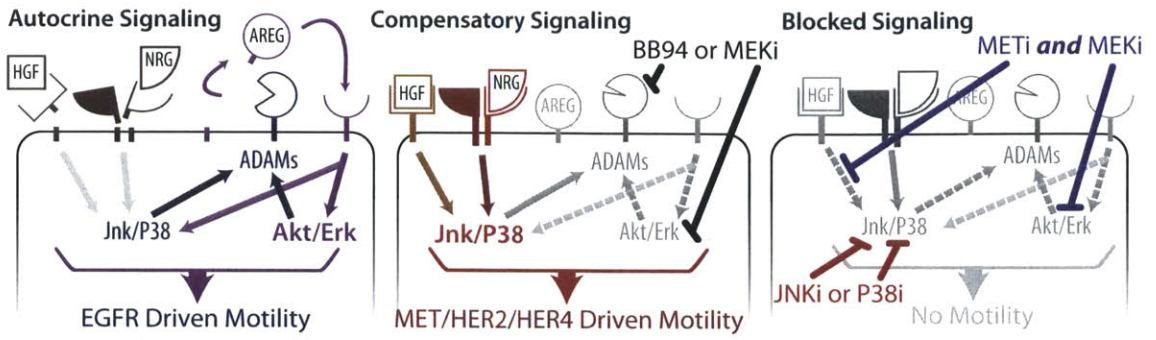

Figure 4-13: **Ectodomain shedding exerts pro- and anti- migratory effects depending on context.** Left: Both direct EGFR stimulation and indirect EGFR autocrine transactivation drive motility in an Erk/Akt dependent manner. High ADAM activity leads to positive EGFR feedback signaling via AREG release, and simultaneously leads to sustained HER2, HER4, and MET shedding. Center: direct inhibition of ADAM activity using protease inhibitors, or indirect ADAM downregulation through Mek inhibition, leads to reduced RTK shedding and subsequently enhanced response to NRG and HGF through Jnk and P38 pathways. Right: Successful inhibition of cellular migration occurs through either direct inhibition of the P38 or Jnk pathway, or through combination kinase inhibition of Mek and MET.

and our clinical evidence associates TIMP-1 with ADAM-10 activity via its relation with MET shedding.

Overall, we have presented an integrative paradigm for analyzing complex networks of protease activities in concert with signaling pathway activities that operate both downstream and upstream of the protease activities via ligand/receptor shedding feedback (Fig. 4-13). We have demonstrated how ectodomain shedding serves as an integrative layer of cellular signaling that critically influences how cells respond to environmental cues and therapeutic interventions. Full appreciation of the many competing roles of ADAM activity be will essential for understanding their role in development and disease, and has wide implications for designing therapeutic strategies.

4.4 Materials & Methods

Full description of the materials used in this work can be found in Sec. C.1. Briefly, the 12Z cell line was generously provided by Anna Starzinski-Powitz (University of

Frankfurt) by way of Steve Palmer (EMD Serono). Throughout the paper, phosphoprotein and supernatant protein measurements were all measured using commercially available ELISAs (often Luminex bead-based) from Bio-Rad, EMD4Biosciences, Millipore, and R&D Systems, with the exception of measurements of MET phosphorylation by western blot (Fig. 9A). Throughout the paper, all experiments were performed at least twice, from separate biological samples, and all reported error bars indicate standard error of the mean (S.E.M.) unless otherwise stated.

For live-cell 3D migration assessment, cells were labeled with a cell-tracker dye (CMPTX; Invitrogen), mixed with DMEM+2.2mg/mL pH-neutralized collagen I (BD Biosciences) at 500,000 cells/mL, placed in a glass-bottom multiwell plate (MatTek; Ashland, MA), polymerized for 30 min. at 37°C, and then overlaid with full serum media overnight. Cells were stimulated four hours before imaging on an environment-controlled Nikon TE2000 microscope. Image stacks of 70 - 3um slices were obtained every 60 min. for 16 hrs. Cells were tracked using Bitplane Imaris (South Windsor, CT). Each of the cell tracks were fit to a random walk model to calculate the random motility coefficient (Kim et al., 2008). For endpoint invasion assays and 3D shedding assays, unlabeled cells were mixed with DMEM+collagen on ice, immediately placed in a standard 96-well tissue culture plate, spun for 5min at 300g, and polymerized for 30min at 37°C. Gels were then bathed in 50ul full serum media containing inhibitors for 60mins, followed by the addition of 50ul full serum media containing growth factors. After 24hr incubation, supernatant was collected, clarified, and frozen for later analysis. Gels were fixed with 1% PFA, stained with YO-PRO-1 (Invitrogen), and imaged at 5X with a CARVII confocal imager (BD) every 3um. Images were analyzed using a modified spot finding algorithm (Santella et al., 2011) in Matlab.

Chapter 5

Conclusions and Future Directions.

This thesis has focused on applying a multivariate, quantitative, and systems-level approach to understanding how extracellular protease networks are regulated in the invasive diseases of endometriosis and breast cancer. In more well-studied and experimentally tractable biological systems, for example mitogenic kinase networks, multivariate and computational approaches have been employed over the last 5-10 years to elucidate "high-level" principals that emerge as an integrated function of interconnected, cooperative, and recursive biochemical reaction networks (Ideker et al., 2003; Liu et al., 2009) . As recent examples, these include an understanding of how complex patterns form during development via cellular communication (Lembong et al., 2009); how pharmacology is affected by networks of biochemical reactions (Kleiman et al., 2011); how complex oscillations may arise among signaling proteins (Shankaran et al., 2009); and how biochemical noise may stabilize population dynamics (Miller et al., 2012b; Acar et al., 2008). In contrast to intracellular signaling, the tools to study metalloproteinase (MP) biology are less well developed, the underlying biological complexity has only recently begun to be appreciated, and very few systems-level studies of extracellular protease biology have been successfully undertaken. Extracellular proteases represent a significant class of drug targets, and many difficulties in clinically targeting them relate to a poor understanding of integrated, network-level biology (Dufour et al., 2013). Although the term "systems-biology" has been applied to MP research in the past (Overall et al., 2006), it has generally been

used to describe -omic studies that examine broad changes in proteolysis across many proteins, but without delving into the biological consequences of feedback, cross-talk, and cooperative interactions. Therefore, a central aim of this thesis has been to bring the types of quantitative, multivariate tools and approaches that have been successful in studying cell-signaling networks to bear on MP biology. The first three chapters of this work are engineering oriented, and describe conceptual frameworks (Chapter 1) and experimental methodologies (Chapters 2-3) as foundational tools with which the science is approached in Chapter 4. Although these methods have proved useful in Chapter 4 and elsewhere (for example, Moss et al., 2011), there remains much room for future development of experimental approaches, in particular.

5.1 Protease activity-based assays.

Because MPs are extensively post-translationally regulated, studies aimed at understanding their regulation inevitably rely upon assessment of protease activity rather than mere expression. Unfortunately, measurement of specific activities has been challenging, and a large portion of this thesis (Chapter 2-3) is aimed at developing methods to address this issue. The PrAMA technique developed in Chapter 2 allows for non-invasive, real-time, multiplexed assessment of proteolytic activities on live cells (Miller et al., 2011; Chen et al., 2012). However, this method has shortcomings that can be addressed through multiple promising avenues.

Live-cell protease activity using FRET-substrates: Live-cell assessment of protease activity is especially critical for observing rapid dynamics while minimally disturbing regulatory factors related to trafficking, non-covalent protein interactions, and post-translational modifications such as phosphorylation or oxidation. PrAMA faces four main limitations in live-cell application: (a) the soluble FRET-substrates for PrAMA do not describe spatially localized protease activity; (b) the FRET-substrates are by nature non-specific, and PrAMA inference accuracy suffers as a result; (c) sensitivity issues limit the temporal resolution and cell-culture applica-

tion, typically requiring dense cultures of highly proteolytic cell populations; and (d) soluble FRET-substrates may not reflect proteolytic activity on membrane-bound endogenous substrates. The following sections provide possible tactics for addressing these four issues.

To generate FRET-substrates that are more specific for particular enzymes of interest, substrates could be optimized using phage display (Chen et al., 2002), synthetic peptide libraries (Lim et al., 2009), or through conjugation to highly-specific affinity reagents like antibodies. FRET-substrate sensitivity could be enhanced with more efficient fluorophores and ratiometric imaging (Preus et al., 2012), or by spatially clustering multiple FRET-substrates for local signal amplification. Spatial clustering could be performed within a polymer (for example similar to RGD clustering; Irvine et al., 2001) or through direct conjugation to micro-particle beads (Carregal-Romero et al., 2013).

Strategies for identifying localized protease activity on live-cells typically involve substrate-immobilization to a matrix scaffold (Packard et al., 2009). One of the most well-known implementations of this method involves conjugating FRET-quenched fluorophores to collagen-I, which then become de-quenched upon collagenolysis. Gel-embedded substrates, including DQ-collagen, have been routinely used for in situ zymography, which can be performed on frozen or gently fixed tissue (Galis et al., 1995). Long incubation times with relatively steady-state tissue allow sufficient proteolysis to occur for signal detection. These approaches provide good sensitivity and localization, but much like the soluble PrAMA substrates, fail to describe the activity of any specific protease. Alternatively, multiple FRET-peptides could be employed in combination using different colored fluorophores. PrAMA inference could then be used to deconvolve the cleavage rates of multiple substrates, and inference algorithms could take localization into consideration. Another promising avenue may be to spatially segregate FRET-substrates within a matrix. Embedding FRET-substrates within spatially confined and identifiable regions of a scaffold would allow for substrate-multiplexing without necessitating multiple fluorophore colors (DeForest et al., 2012). Moreover, spatial segregation could be used to enhance protease specificity. Substrates could

be buried within a matrix to segregate cell-surface proteases from secreted. Physiochemical barriers such as restricted matrix pore-size, pH, or hydrophobicity could selectively modulate transport and catalytic efficiencies of various proteases. Overall, FRET-substrate immobilization represents a promising approach for non-invasively detecting protease activity on live-cells, yet faces specificity and sensitivity hurtles.

Membrane-bound substrate proteolysis: Direct examination of endogenous, membrane-bound substrate proteolysis is generally a more biologically relevant indicator of protease activity compared to observing FRET-substrate cleavage, insofar as the endogenous substrates are themselves regulated and can affect downstream biology. Chapter 4 of this thesis, along with other reports (Herrlich et al., 2008; Dang et al., 2011; Inoue et al., 2012), have demonstrated how directly monitoring membrane-bound substrate cleavage through detection of soluble-fragment accumulation or using cell-surface immunostaining can be a valuable metric of protease activity. However, data in Chapter 4 and elsewhere (Dang et al., 2011) also emphasize that membrane-bound substrates exhibit the same issues of non-specificity as the soluble FRET-substrates: endogenous or membrane-bound substrate shedding often cannot be uniquely attributed to a specific protease, and may also be a function of substrate-specific modifications such as trafficking. In contrast, soluble FRET-substrates provide the relatively non-invasive (that is, not requiring transgenic substrate expression) examination of active protease concentration in a manner that is essentially independent of endogenous substrate regulation. Membrane-bound substrates are especially useful for understanding substrate-specific regulatory pathways. Moreover, they enable single-cell cleavage detection, and therefore support shRNA library screening and investigations of intercellular heterogeneity. Future directions will likely focus on simultaneously measuring proteolysis of multiple membrane-bound substrates to understand their specific regulation and integrated downstream effects.

Assessing protease activity in fixed or lysed cells and tissues: Both sensitivity and specificity can be improved in more invasive experimental approaches.

Protease inhibitors generally target the enzyme active site and can be tagged to function as activity-based probes (ABPs) for a variety of applications, including fluorescence microscopy for subcellular resolution of active enzyme concentrations (Kcow et al., 2012). In general, however, these probes are non-specific among closely related enzymes, and future directions for this technology involve improvements to probe specificity. Antibody-based approaches have been demonstrated as promising for specifically binding active protease (Tape et al., 2011), and small molecule protease inhibitors have largely improved in specificity over the last decade (Devy et al., 2011). Future work may involve leveraging antibody specificity with active-site probes using fluorescence co-localization or proximity ligation. ABPs can be used to purify active proteases for downstream analysis using mass spectrometry or various immunoassays. Future directions for this approach will likely involve improved ABP reagents for reduced cross-reactivity and greater purification capabilities for MS/MS applications (Sieber et al., 2006). Although not specific to proteases themselves, ongoing work continues to improve the breadth, quantitation, reproducibility, and throughput of MS/MS based proteomic methods. For more targeted assays, antibodies have been successfully used to immunopurify proteases for subsequent activity-based analysis with either ABPs or FRET-substrates, and future directions for this approach will likely focus on enhanced throughput, sensitivity, and multiplexing (for instance by using microfluidics or suspension array technology).

Assessing protease activity in microfluidic platforms: Chapter 3 of this thesis describes how the large sample-requirement of PrAMA can be overcome using droplet-based microfluidics; however, several promising improvements and alternatives to this solution exist. One of the chief limitations to the microfluidic PrAMA lies in the small size of its droplet library. In Chapter 3 we construct droplet libraries comprising up to eleven uniquely identifiable droplet-types based on their optical bar code. This size is currently restricted due to a limited number of optically-discernible dyes and dye-concentrations, along with the manual and serial construction of each type of droplet. Therefore, future directions will likely be aimed at generating droplet libraries in an

automated, highly parallelized fashion. For example, libraries could be pre-assembled on a separate device that simultaneously loads multiple reaction mixtures from a microtiter plate (for example, as proposed in Guo et al., 2012), or on a single integrated chip with the biological-sample injector. Regardless of the strategy for droplet generation, large libraries will require sophisticated droplet barcoding schemes. Integrating droplet library generation on-chip with sample-injection would support confining the droplet library to single-file within a microfluidic channel, subsequently allowing for positionally encoded droplets. As another option, different types of droplets could be identified by encapsulated microparticles (for instance, Pregibon et al., 2007), which has the potential to support the unique identification of over 1000 types of droplets. Ultimately, a larger droplet library would allow for additional protease substrates and inhibitors to be combined at varying concentrations for a single biological sample. As a result, more protease activities could be measured with greater accuracy; inhibitor or drug combinations could be tested for synergistic effect; and dose-response curves for various inhibitors could be calculated for a given biological sample. Alternative microfluidic platforms have begun to be commercialized for small-volume combinatorial reactions, and these show promise (for instance, platforms from Fluidigm and Wafergen). However, the droplet-based approach exhibits the greatest promise for monitoring rapid reaction kinetics and precisely controlling fluid mixing.

5.2 Network-level aspects of protease regulation.

In addition to developing protease activity assays, this thesis focuses on applying them to understand and target sheddase activity in disease. This work puts forth, as a proof-of-principle, how a systems-level "cue-signal-response" approach to understanding MP regulation can lead to novel insights and therapeutic strategies. Many future directions exist for the application of systems-biology tools to study and target extracellular proteases, and we highlight several in the following section.

The biological chemistry of protease regulation: At a fundamental level, work continues to be done in mapping out the known web of biochemical interactions among proteases, their inhibitors, and their endogenous substrates. Although the last decade has seen progress in this regard, the catalog of basic enzyme-substrate relationships still remains incomplete. Furthermore, biochemical interactions have increasingly become acknowledged as complex: cleavage reactions can inactivate, modify, or activate their substrates in unexpected manners. From a forward-engineering systems-biology perspective, a basic registry of biochemical reactions among proteases needs to first be determined and understood before mechanistic models of protease networks can properly be constructed. On the other hand, the experiments needed to understand nuanced biochemical relationships among proteases can be extremely time- and resource- intensive, and reverse-engineering methods from systems-biology are likely to be useful tools for identifying critical components in the protease network to study. Network-level modeling approaches may be especially important for understanding and identifying higher-order protease interactions that involve multiple protein components or pathways.

Beyond identifying individual biological components, network-level approaches to protease biology may be particularly useful in elucidating mechanisms of protease regulation. Multiple biological processes cooperatively interact to influence protease activity, including trafficking, post-translational modification (for example phosphorylation), non-covalent dimerization, and substrate availability. Analogous to studies exploring growth-factor receptor signaling and trafficking (Lauffenburger et al., 1996), quantitative computational modeling will likely be helpful in assessing how different biochemical processes related to proteases interact to impact overall behavior. In many cases, the basic biochemical components and governing equations that describe certain key proteases may be well known, and yet their functional behavior as an integrated network of biochemical reactions has not been well understood. As an example of this, network-biology based approaches are well suited to help explain, for instance, the role of stoichiometry among multiple interacting proteases and their endogenous inhibitors, which generally have overlapping enzymatic or inhibitory specificities, re-

spectively. Mathematical modeling may be especially critical to understanding proteases and their interaction with kinase signaling networks, where complex dynamics and feedback mechanisms play a significant role. This thesis explores the role of ligand and receptor shedding feedback on kinase signaling, and future studies may investigate these relationships in much greater detail, with different proteases and substrates, and in different disease models. Overall, network-level approaches show promise as tools to elucidate the inherently interconnected and complex web of extracellular protease interactions.

The in vivo protease network: Extracellular protease regulation becomes especially complex in the in vivo context, where intercellular communication, reaction/diffusion, and mechanical considerations can play significant roles. Extracellular proteases can be regulated by oxidative stress (Katsu et al., 2010), mechanical properties of surrounding extracellular matrix (Leong et al., 2011), growth-factor/cytokine signaling (Chapter 4 of this thesis), and by cycling hormones including estrogen and progesterone (Curry et al., 2003). In turn, proteases are widely recognized as directly and concomitantly contributing to all of these regulatory factors. As one example feedback relationship, MP activities are sensitive to oxidative stress and hypoxia, and at the same time mediate signaling and cell migration processes related to angiogenesis and ECM diffusivity. The integrated consequences of TIMP and MMP interactions grow increasingly complex as reaction/diffusion consideration are taken into account, especially given many proteases and their inhibitors can be found at high circulating concentrations. Extracellular proteases and inhibitors have been found capable of directly binding ECM components, adding yet another layer of complexity into understanding in vivo reaction/diffusion kinetics. In this work, we find peritoneal levels of proteases, inhibitors, and protease substrates to all be highly covariate and jointly implicated in endometriosis. Future work may investigate how various cell populations contribute to the complex mix of proteases and inhibitors that systemically accumulates with disease, and how reaction/diffusion and intercellular feedback mechanisms integrate to affect overall disease progression. We anticipate that mul-

tivariate and quantitative approaches such as the cue-signal-response paradigm are likely to be valuable in interpreting and prioritizing in vivo observations of this complex biology.

Targeting the protease network in disease: Although this work chiefly investigates breast cancer and endometriosis, extracellular proteases and the signaling pathways with which they participate have been implicated in many cancers and inflammatory diseases. In this work we examine how various kinase signaling pathways affect sheddase activity, and many of the inhibitors we tested represent major drug targets in a variety of settings. EGFR inhibitors, for example, are clinically approved for non-small-cell lung cancer with ongoing clinical trials for other indications. Mek, PI3K, and p38 inhibitors are currently in clinical trials for multiple advanced malignancies. P38 inhibitors in particular are being investigated for various inflammatory diseases, most notably rheumatoid arthritis. Furthermore, many sheddase substrates are attractive drug targets for these diseases, and some substrates have clinically approved therapies for various indications (for example, HER2 is targeted in HER2+ breast cancer).

This thesis presents a fairly complex image of sheddase regulation characterized by cross-talk and feedback relationships; nevertheless, several clear conclusions can be drawn regarding therapeutic strategy. Our results from the peritoneal fluid analysis suggest that endometriosis is associated with an imbalance in multiple closely related protease substrates, and emphasize that the disease is heterogeneous at the molecular level from patient-to-patient. Nevertheless, multivariate and quantitative methods were valuable in interpreting this heterogeneity and finding interpretable (and therapeutically actionable) features of disease. In fact, multiplexed measurements were required in this case to identify statistically significant patterns in the clinical data. We anticipate that similar combination of activity-based analytical measurements with multiplexed assessment of downstream activity markers (in this case, substrate shedding) will be applicable to a broad range of pathologies, and will be necessary for identifying the most relevant patients to treat with a particular drug. Future work

may also investigate using these approaches in more readily obtained biological fluids, such as plasma, urine, or by using molecular measurements made through clinical imaging modalities.

This work focuses most heavily on the feedback relationship between kinase inhibition and RTK shedding in endometriosis, and future work may investigate the ubiquity of this mechanism in other diseases where drugs have been developed to target signaling pathways that regulate sheddase activity. For instance, Mek inhibitors have been particularly well-studied in malignant melanoma, where several sheddase substrate receptors (including Tyro-3) are known to be critical (Zhu et al., 2009). Therefore, it would be interesting to investigate whether Mek inhibition leads to a compensatory upregulation of Tyro-3 in melanoma, and to test whether combination Mek/Tyro inhibition synergistically blocks disease progression. We also found a novel role for p38 and Jnk signaling pathways in the context of compensatory RTK signaling. Inhibitors targeting these pathways were successful in broadly reducing cell migration in multiple cancer cell lines, although we did not detect a significant impact on cell growth. Therefore, future studies may investigate whether the Jnk and p38 pathways could be targeted in combination with more cytotoxic drugs for preventing cell-migration-dependent cancer metastasis.

Observed cross-talk between ADAM-10 and ADAM-17, coupled with their catalytic promiscuity, make these sheddases problematic drug targets. However, recent efforts at developing pharmacological protease-modulating drugs rather than protease inhibitors have shown promise, particularly with gamma-secretase as it relates to Alzheimer's disease (Kukar et al., 2008). These small molecule agents function by binding substrate directly, and our results suggest that an analogous approach could be applied to ADAM-10 and ADAM-17 substrates, where RTK shedding but not ligand shedding may be desired. In fact, a recently described anti-MET antibody has been observed to inhibit signaling by, at least in part, enhancing MET shedding (Schelter et al., 2010). Importantly, a challenge to all approaches directly targeting sheddase substrates lies in the fact that multiple sheddase substrates may compensate for each other. For example in this work we not only find MET, but also HER2 and

HER4 to be implicated in compensatory signaling. Next-generation targeted therapies are actively being developed to target multiple related signaling pathways or RTKs (McDonagh et al., 2012), and development of these drugs may well be served by consideration of sheddase-mediated feedback. Intracellular endogenous modulators of sheddase activity are actively being sought after as another way to target multiple sheddase substrates, and reports show that various PKC isoforms can impact the cleavage of closely related sheddase-substrates in a biased manner (Dang et al., 2011; Kveiborg et al., 2011). However, truly selective pathways of sheddase regulation that clearly distinguish between EGF-ligand and RTK shedding, for example, have yet to be identified.

Overall, this work attempts to develop and use tools from systems-biology to gain enhanced understanding of protease regulation from a network-level perspective. Our hope is that the results from this study have demonstrated the potential value in using quantitative, multivariate methods to understand cross-talk and feedback mechanisms inherent to protease biology; and that the tools and concepts presented herein may help guide future drug-development efforts.

Appendix A

Supporting Information for Chapter 2.

A.1 Comments regarding substrate depletion.

Substrate depletion often occurs in our experiments involving purified recombinant enzyme. For example, roughly 70% of the initial substrate is degraded in the reaction shown in Fig. 2-2A. Accounting for substrate depletion extends the quantitative range of the assay and has no negative impact on inference when actual depletion is negligible (Fig. 2-2B). Lastly, substrates can become depleted in live-cell applications. For example, basal proteolytic activity in mouse embryonic fibroblasts results in roughly 25% of substrate 6 being degraded within three hours (Fig. A-9). After three hours, raw fluorescence for that sample is ~6000 FLU. Peak fluorescence of the corresponding positive control is roughly ~25000 FLU (data not shown).

A.2 Abbreviations

ABP, activity based probe;

ADAM, a disintegrin and metalloproteinase;

B, background signal;

DMEM, Dulbecco's Modified Eagle Medium;

DMSO, dimethyl sulfoxide;

$C_{i,j}$, catalytic efficiency for the ith substrate and jth enzyme;

C, catalytic efficiency parameter matrix;

Cha, cyclohexylalanyl;

Dab, Dabcyl, 4-(4-dimethylaminophenylazo)benzoyl;

$[E]$, enzyme concentration;

E, vector of enzyme activities;

EGF, epidermal growth factor;

F_0, peak fluorescence from positive control;

F_{obs}, observed fluorescence from product formation;

F_p, fluorescence from product formation;

Fam, 5-carboxyfluorescein;

FRET, fluorescence resonance energy transfer;

GABA, γ-aminobutyric acid;

Homophe, homophenylalanyl;

IM, ionomycin;

k_{cat}, turnover number;

k_{cat}/K_m, catalytic efficiency;

k_d, photobleaching decay constant;

K_m, Michaelis-Menten constant;

M-M, Michaelis-Menten;

MEBM, Mammary Epithelial Basal Medium;

MMP, matrix metalloproteinase;

MP, metalloproteinase;

NorVal, norvaline;

PrAMA, Proteolytic Activity Matrix Analysis;

R_M, model covariance error matrix;

R_M^r, relative model covariance error matrix;

R_D, data uncertainty covariance matrix;

$[S]_0$, initial substrate concentration;

$[S]$, substrate;

σ_T, significance of inference threshold;

T_0, lag time;

TIMP, tissue inhibitor of metalloproteinase;

TNFα, tumor necrosis factor-alpha;

V_0, initial rate of substrate cleavage;

$\mathbf{V_0}$, vector of V_0's for all substrates;

$\mathbf{V_0^s}$, bootstrapping sample ensemble of multiple $\mathbf{V_0}$.

Ref. #	Cat. #	Substrate Sequence
1	011	Dab-Gly-Pro-Leu-Gly-Met- ...-Arg-Gly-Lys(5-FAM)-NH2
2	014	Dab-Glu-His-Ala-Asp-Leu- ...-Leu-Ala-Val-Val-Ala-Lys(5-FAM)-NH2
3	021	Dab-Val-Pro-Val-Asn-Norleu- ...-Thr-Val-Lys(5-FAM)-NH2
4	015	Dab-Val-Asp-Leu-Phe-Tyr- ...-Leu-Gln-Gln-Pro-Lys(5-FAM)-NH2
5	008	Dab-Pro-Cha-Gly-Cys(Mc)His-Ala- ...-Lys(5-FAM)-NH2
6	022	Dab-Leu-Arg-Ala-Glu-Gln- ...-Gln-Arg-Leu-Lys-Ser-Lys(5-FAM)-NH2
7	005	Dab-Leu-Ala-Gln-Ala-Homophe- ...-Arg-Ser-Lys(5-FAM)-NH2
8	017	Dab-Ala-Pro-Arg-Trp-Ile- ...-Gln-Asp-Lys(5-FAM)-NH2
9	010	Dab-Ser-Pro-Leu-Ala-Gln- ...-Ala-Val-Arg-Ser-Ser-Lys(5-FAM)-NH2
10	052	Dab-Ala-Pro-Phe-Glu-Met- ...-Ser-Ala-Lys(5-FAM)-NH2
11	016	Dab-Ser-Asn-Leu-Ala-Tyr- ...-Tyr-Thr-Ala-Lys(5-FAM)-NH2
12	059	Dab-Ala-Pro-Arg-Trp-Leu- ...-Thr-Thr-Ala-Lys(5-FLU)-NH2
13	053	Dab-Ala-Pro-Phe-Glu-Phe- ...-Ser-Ala-Cys(5-FLU)-NH2
14	012	Dab-Val-Pro-Phe-Glu-Phe- ...-Thr-Val-Lys(5-FAM)-NH2
15	013	Dab-His-Gly-Asp-Gln-Met- ...-Ala-Gln-Lys-Ser-Lys(5-FAM)-NH2
16	020	Dab-Val-Pro-Thr-Trp-Ile- ...-Gln-Asp-Lys(5-FAM)-NH2
17	200	Dab-GABA-Arg-Pro-Lys-Pro- ...-Val-Glu-NorVal-Ala-Arg-Cys(5-FLU)-Gly-CONH2
18	201	Dab-GABA-Pro-Gln-Gly-Leu- ...-Cys(5-FLU)-Ala-Lys-CONH2

Table A.1: **Synthetic polypeptide FRET-substrate sequences.** Reference numbers denote substrate indices used throughout the chapter.

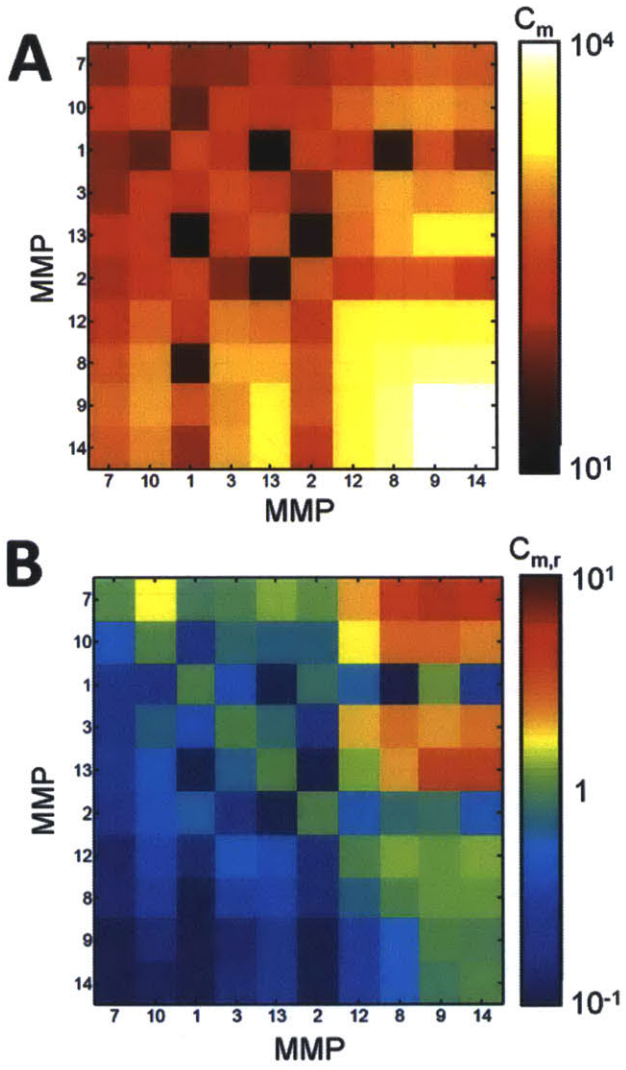

Figure A-1: **R_M transformed to emphasize additive rather than multiplicative error.** Columns of the parameter matrix C were normalized by their Euclidean norm prior to calculation of R_M.

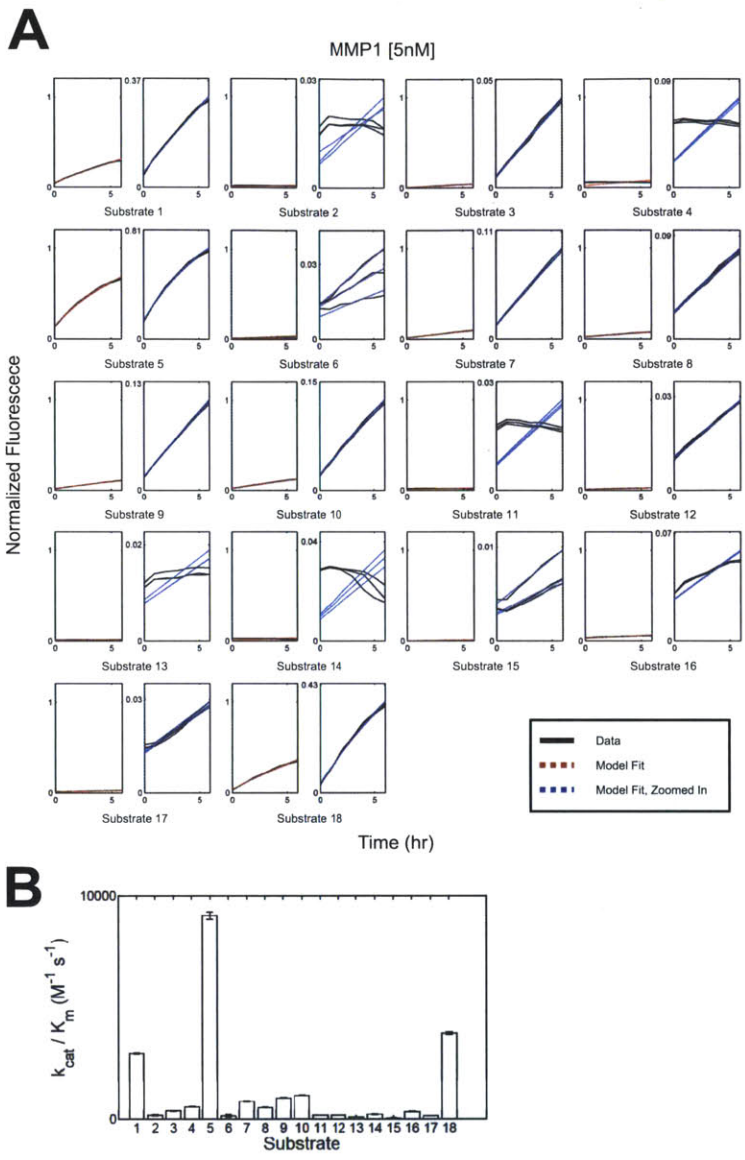

Figure A-2: **Example Cleavage Signature of MMP1.** kcat/Km values were determined using MMP1 at 0.4nM and individual substrates at 10uM in MMP minimal buffer (see text). (A) Time-lapse fluorimetry data, normalized to the positive control signal, $F_{0,S}$. Blue and red plots are paired, such that blue plots show the identical data but at a zoomed in scale on the y-axis. Both blue and red lines indicate fit to the non-linear kinetic model. Individual lines indicate technical replicates. (B) kcat/Km values are inferred from each of the time-courses shown in A. Error bars indicate standard deviation.

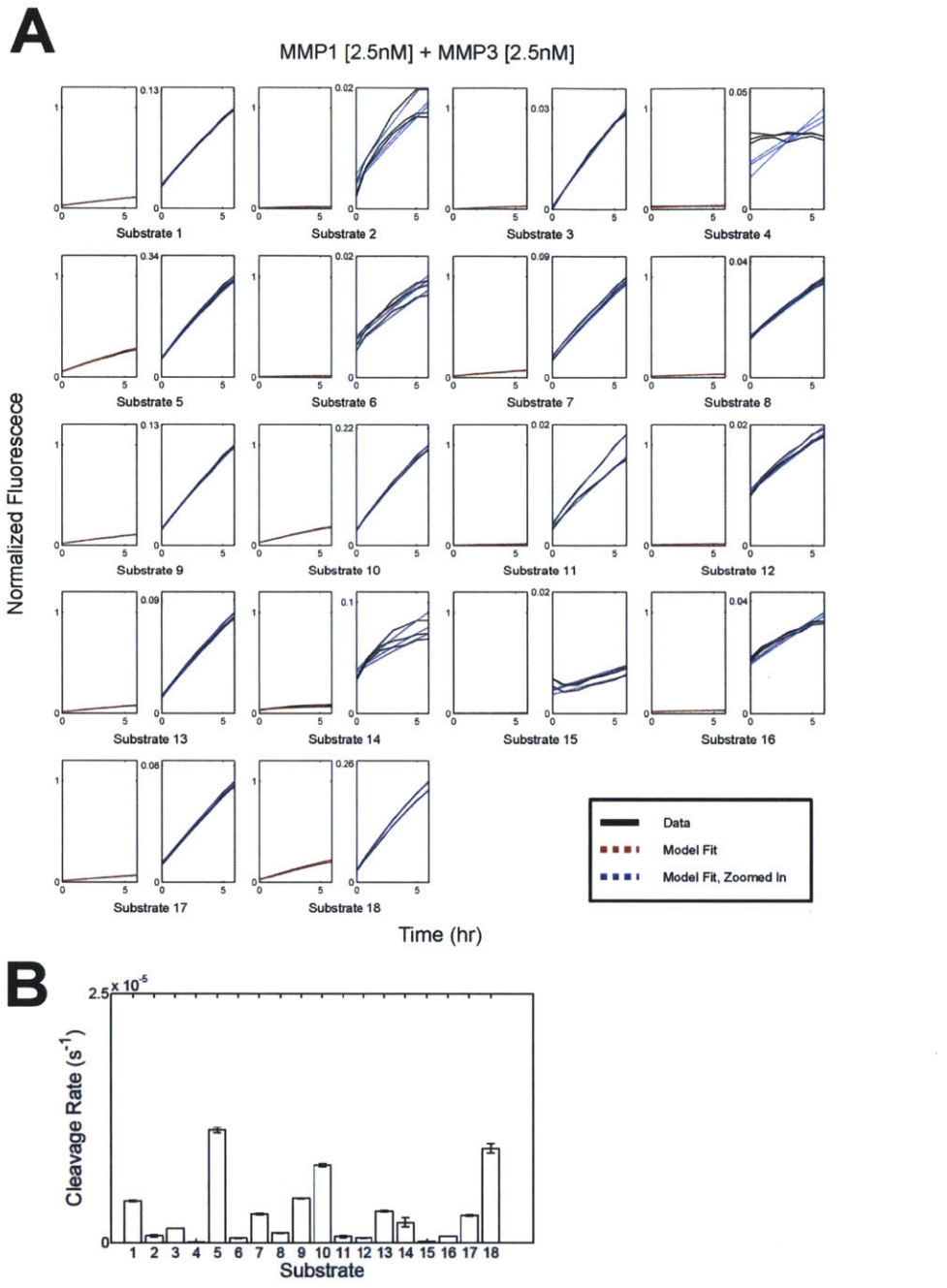

Figure A-3: **Example Cleavage Signature of MMP mixture.** kcat/Km values were determined using 0.2nM MMP1, 0.4nM MMP3, and individual substrates at 10uM in MMP minimal buffer (see text). (A) Time-lapse fluorimetry data, normalized to the positive control signal, $F_{0,S}$. Blue and red plots are paired, such that blue plots show the identical data but at a zoomed in scale on the y-axis. Both blue and red lines indicate fit to the non-linear kinetic model. Individual lines indicate technical replicates. (B) Cleavage rates are inferred from each of the time-courses shown in A. Error bars indicate standard deviation.

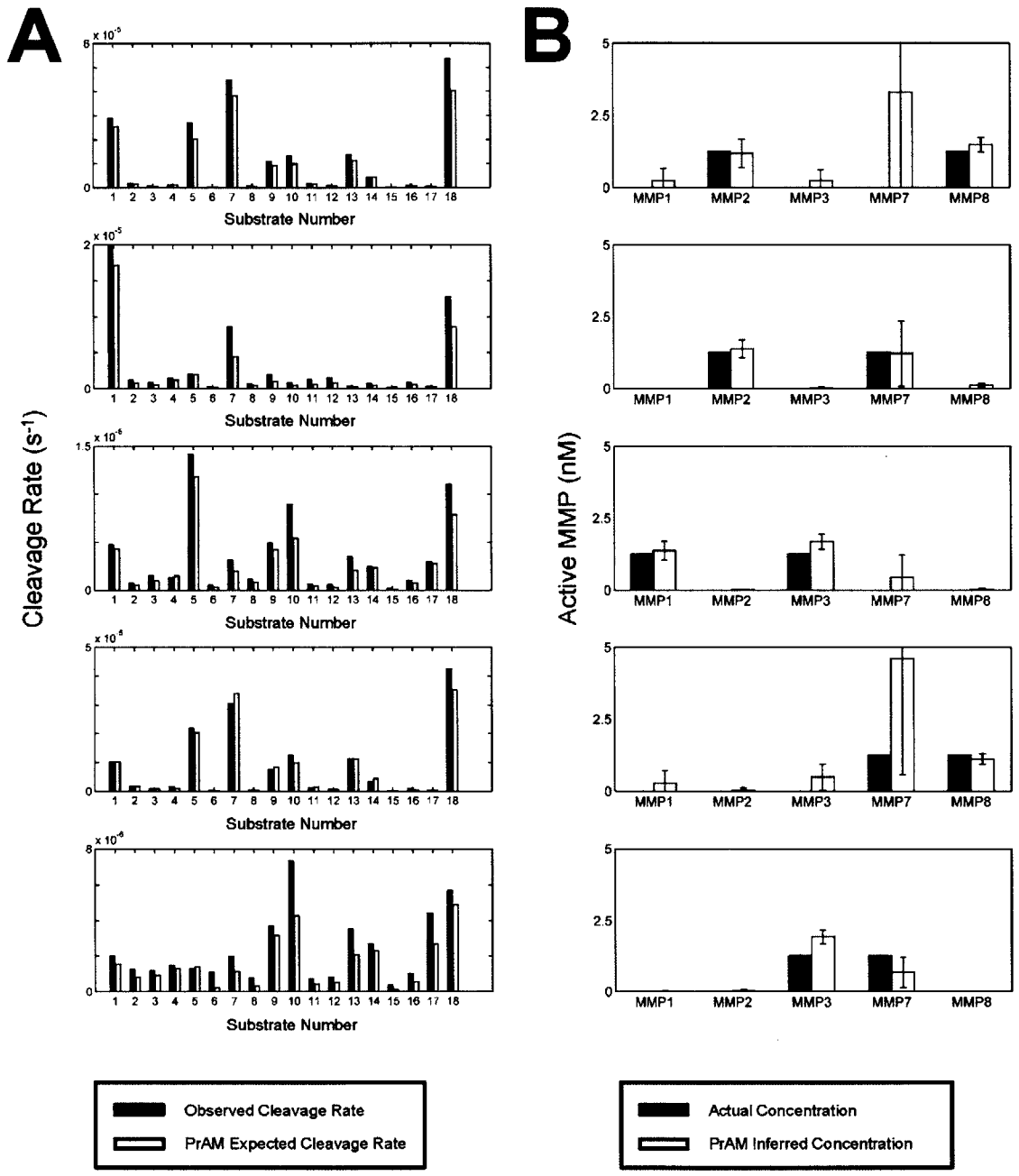

Figure A-4: **Summarized Cleavage Signatures of MMP mixtures and PrAM Inference Results.** (A) Cleavage rates were observed for various mixtures of MMPs (black, Fig. A-3) and compared to rates expected by PrAMA based on the known mixture composition and observed cleavage rates for single enzyme mixtures (white, Fig. A-2). (B) Actual mixture compositions (black) and PrAMA inference results (white) correspond to the cleavage signatures in A. Error bars indicate std. dev. of inference, using 20% synthetic sampling error.

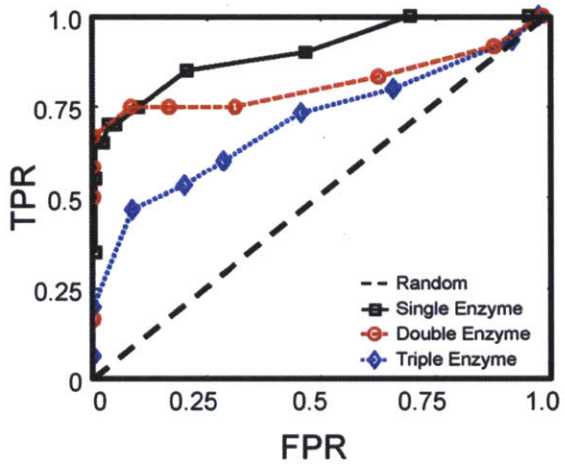

Figure A-5: **PrAMA inference using alternative bootstrapping scheme.** PrAMA inference was conducted on enzyme mixtures containing MMPs 1–8, corresponding to Fig. 2-8, but using a modified bootstrapping scheme that more directly considers uncertainty in the parameter matrix C. Max accuracy is roughly 95% for single and double enzyme mixtures. For triple enzyme mixtures, max accuracy is roughly 80%.

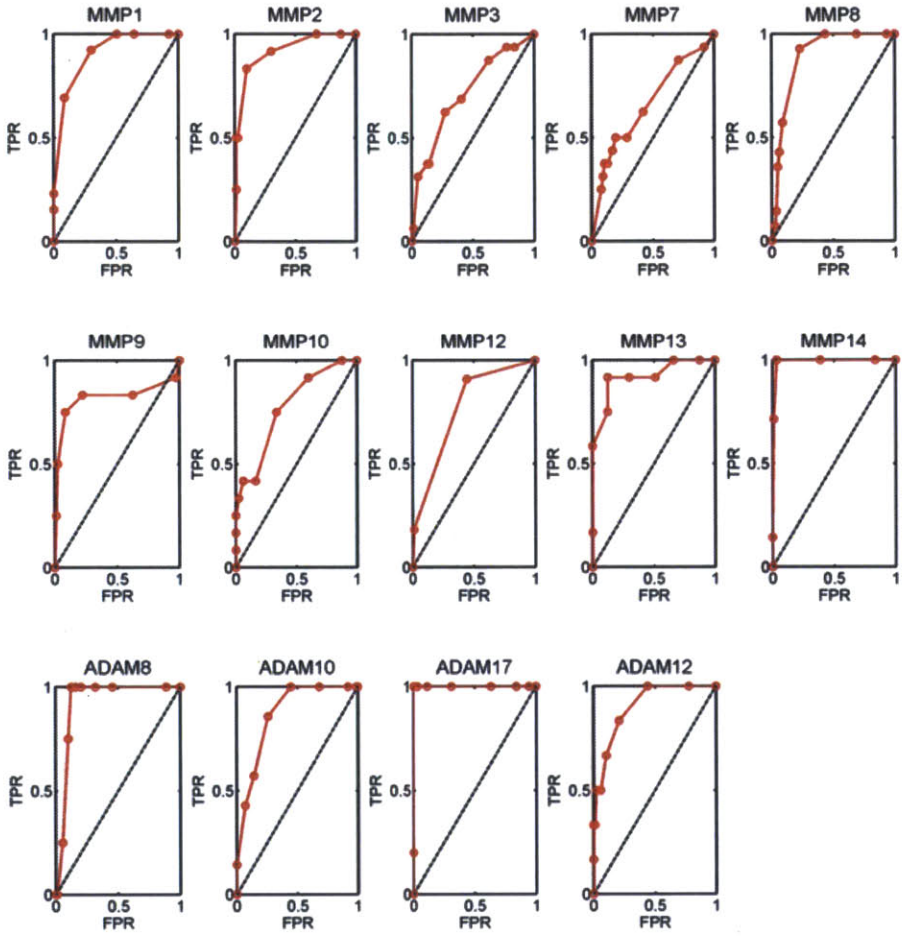

Figure A-6: **ROC Curves describing inference accuracies of individual enzymes.** PrAMA inference corresponds to results shown in Fig. 2-9C. True positive rate (TPR) and false positive rate (FPR) refer to inference results for each individual enzyme, for all mixtures tested and using a 16 substrate x 14 enzyme parameter matrix C (see Fig. 2-8A).

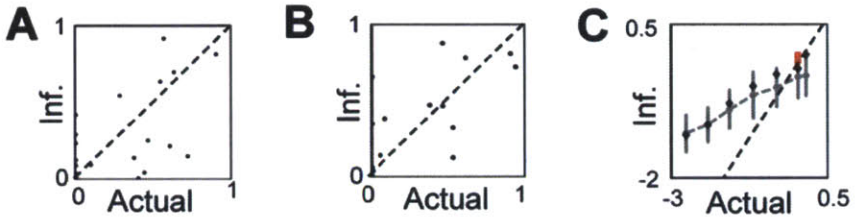

Figure A-7: **PrAMA quantitative inference.** PrAMA inference was performed on all double enzyme mixtures involving MMPs 1-8 (A) and ADAMs 10 & 17 (B). These plots show actual concentrations vs. PrAM inferred concentrations, where enzyme levels were first normalized such that total enzyme concentration in a given mixture sums to unity. R^2 values for (A) and (B) are both roughly 0.5. (C) Black diamonds indicate PrAM inference (ordinate) of MMP7 at various actual concentrations (abscissa), based on the cleavage signature V_0 from substrates 1-16. Red points show PrAM inference of MMP7 in conditioned media from breast cancer cells. Axes are $\log_1 0$-transformed and scaled such that the maximum actual concentration = 1. All MMPs were considered in the parameter matrix C. The grey line indicates inferred $V_{0,i}$, averaged over all 16 substrates, with standard deviation shown by error bars. The dashed line indicates linear inference.

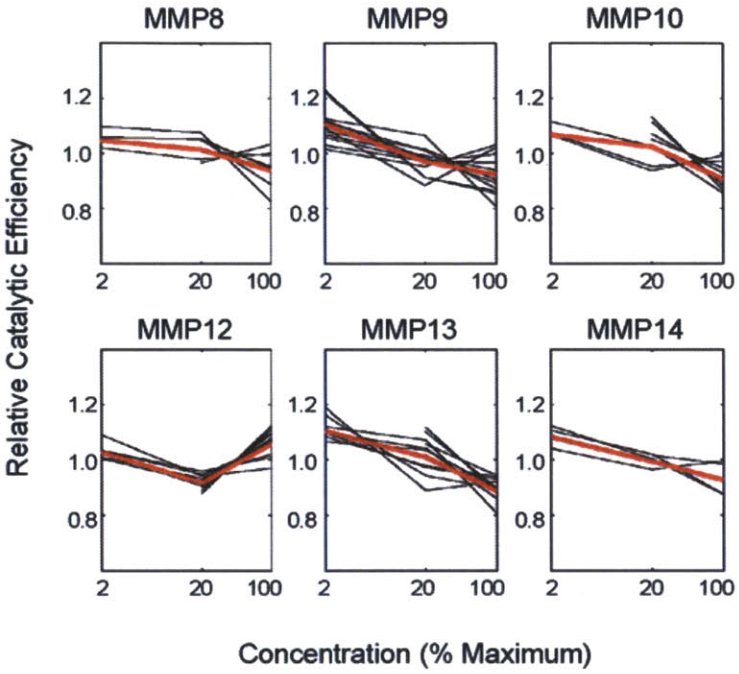

Figure A-8: **Concentration effects on enzyme catalytic efficiency.** Substrate catalytic efficiencies $C_{i,j}$ were calculated at three different concentrations of MMPs (see Fig. 2-4) using 16 substrates. We define relative catalytic efficiency as the log-transformed $C_{i,j}$ divided by the average log-transformed $C_{i,j}$ from all MMP concentrations. Each black line corresponds to the relative catalytic efficiencies observed from individual substrates, and the red line denotes the average over all substrates. Experiments with the lowest two concentrations used MMP buffer, and the highest concentration experiments used MEBM. We only considered $C_{i,j}$ measurements detected significantly above background signal.

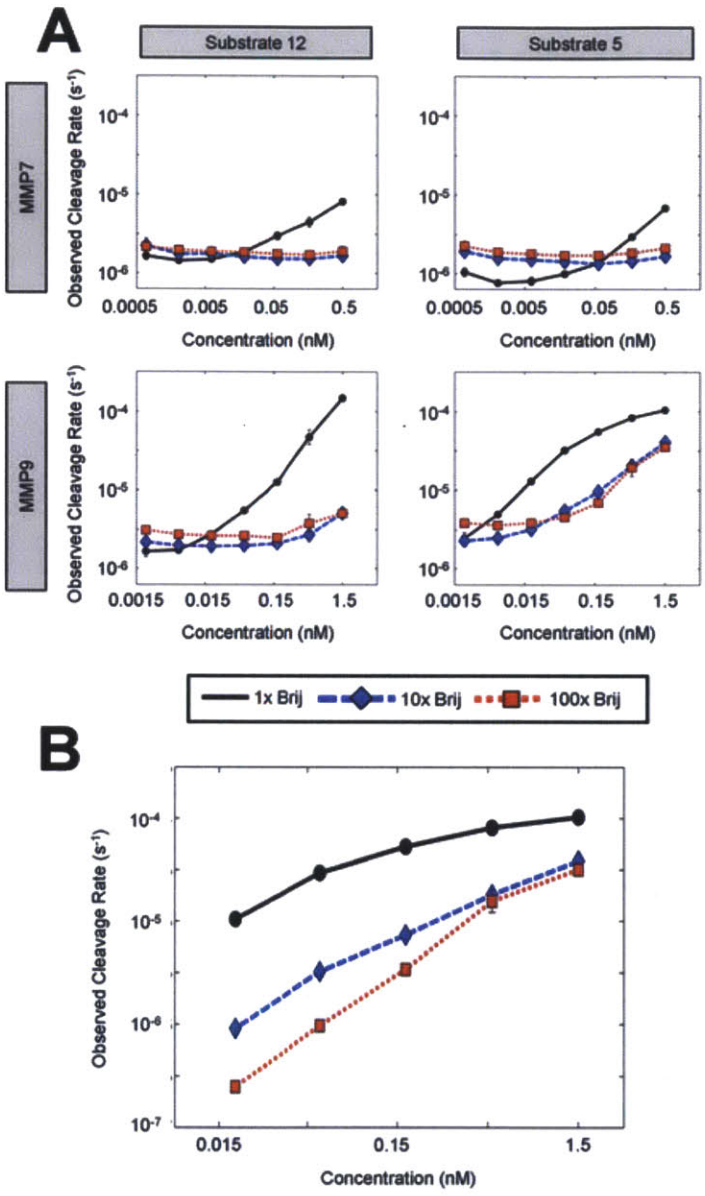

Figure A-9: **The impact of Brij-35 on MMP proteolytic activity.** MMP7 and MMP9 proteolytic activities were observed in MMP buffer with increasing amounts of Brij-35 (10x and 100x the concentration for MMP buffer). (A) Cleavage rates (V_0 / [S]) were calculated from time-lapse fluorimetry data. To emphasize Brij effects on background proteolysis, background cleavage rates were not subtracted in this case. (B) MMP9 cleavage of substrate 5, corresponding to the lower right panel of A. In this case, the background cleavage rates for each Brij concentration were subtracted from the total observed rates. The slopes for the 1x, 10x, and 100x Brij lines on log-log axes are 0.50, 0.82, and 1.0. A slope of 1.0 indicates assay linearity.

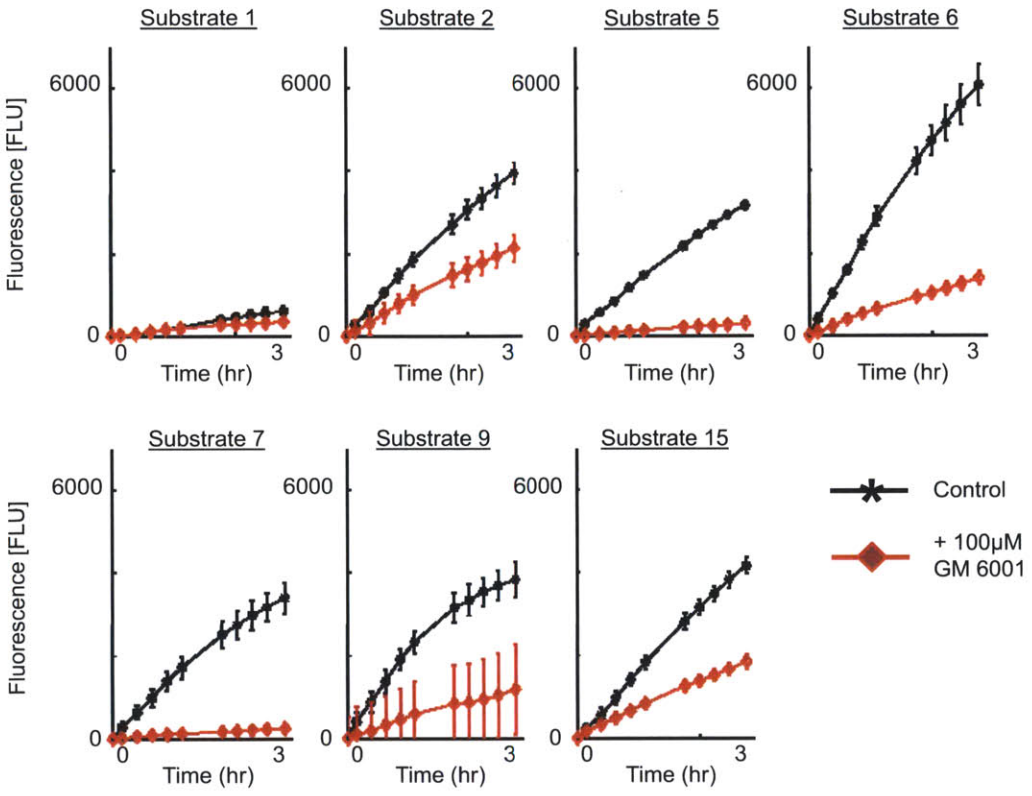

Figure A-10: **Impact of GM6001 treatment on observed protease activity.** WT MEFs were seeded at 30,000 cells per well in a 384-well plate. The following day, media was replaced with fresh media +/- 100uM GM6001. After an hour incubation with the inhibitor, cells were treated with fresh 100uM GM6001 and one of seven FRET substrates, used at a final concentration of 10uM. Cells were imaged for 3hrs at 37°C using a plate reader. Error bars indicate standard deviation of biological quadruplicates.

Appendix B

Supporting Information for Chapter 3.

B.1 Device fabrication and operation

The device was fabricated as a polydimethylsiloxane (PDMS, Sylgard 184, Dow Corning Inc., Midland, MI) chip bonded to a PDMS-coated glass slide. The fabrication of the microchip includes the following steps: (1) master fabrication, (2) PDMS pouring/curing step, (3) spin-coating PDMS on the surface of a glass slide (50mm x 75mm) and (4) irreversible bonding of the PDMS microchannel and the PDMS coated slide via plasma bonding. (5) The fabrications of electrodes in the picoinjectors.

SU8 photoresist (SU8-2050, MicroChem Inc., Newton, MA) was patterned on a silicon wafer to build a positive master. The positive master mold for the device contains channels that are 45um tall. The SU-8 master was further treated with a hexamethyldisilane (Sigma-Aldrich, St. Louis, MO) for 1 hour to prevent adhesion with PDMS after molding. The hexamethyldisilane solution was evaporated and deposited on the master in a desiccator with a ~5 psi vacuum. In the second step, PDMS was degassed in a desiccator with a ~5psi vacuum for 1hr and poured on the master mold. The mold was cured in an oven at 65°C for 6hrs, and then the PDMS layer was peeled off from the silicon master. Holes were punched through the end of the channels using a metal syringe needle with an outer diameter of 1/16in.

(Hamilton Co., Reno, NV). To form the hydrophobic surface for making the droplet generator, a glass slide was coated with a layer of PDMS. To obtain a thin coating, PDMS was diluted by adding Hexane (Sigma, 1:1 mix) and was coated on a glass slide by using a spin coater at 1800rpm. The coated glass was then put into oven at 65°C overnight. After plasma bonding, the device was put into oven at 65°C for more than 20hrs for forming strong bonding and completely hydrophobic surfaces. To make the electrodes in the devices, the empty microchannels in the shape of the electrodes were first constructed1. We heated the devices and injected a low melting-point liquid solder into the empty channels. After cooling the devices, the solder was solidified to form the electrodes embedded in the microfluidic devices.

For device operation, one aqueous and two oil streams were introduced into a droplet generator with co-flow channel geometry using syringe pumps (Harvard, PHD2000). The oil flow rate (10 uL/min) roughly matched aqueous flow rates (5 uL/min) used to form the droplets (~30 pL/droplet with a generation rate of ~4 kHz). For pico-injection, droplets passed through a narrow channel of size similar to the droplet diameter (50 um) at flow rate ~0.03 uL/min. Oil was added from a side channel at ~0.05 uL/min, maintaining the spacing between the drops for synchronization with the pico-injector. To tune picoinjector function, experiments using dye-labeled samples (Alexa-546) and droplets (Alexa-405) ensured a 1:1 mixing ratio during droplet pico-injection. After pico-injection, the diameter of a droplet changes from ~35um (30pL) to ~50um (60pL), and the intensity of indicator dye decreases by roughly 50% after substracting off the background fluorescence. Judging by the changes in observed droplet volume and indicator dye intensity, we conclude the loading efficiency to be roughly 1:1 (droplet : sample). After this characterization, the injection flow rate was controlled as ~0.03uL/min and the DC voltage applied was ~80V with frequency of 1kHz. Time-lapse fluorescence microscopy was performed using a triple-band filter (Semrock) with emitter wavelengths of 457nm, 530nm, and 628nm.

B.2 General reagents & procedures

FRET-substrates (Biozyme, Inc; Apex, NC) contained the following polypeptide sequences: S1: Dabcyl - Gly - Pro - Leu - Gly - Met - Arg - Gly - Lys (5-FAM) - NH2; S2: Dabcyl - Pro - Cha - Gly - Cys (Me) His - Ala - Lys (5-FAM) - NH2; S3: Dabcyl - Leu - Arg - Ala - Glu - Gln - Gln - Arg - Leu - Lys - Ser - Lys (5-FAM) - NH2; and S4: Dabcyl - Ser - Pro - Leu - Ala - Gln - Ala - Val - Arg - Ser - Ser - Lys (5-FAM) - NH2 2. Materials and buffers for the purified enzyme experiments, along with all microtiter-plate assays, were performed as previously described unless otherwise stated. Reaction conditions for the nine-component droplet library included the following in PBS, when appropriate: 10 uM substrate, 10 uM BB94 (Tocris Bioscience; Ellisville, MO); 10 uM pro-ADAM-9 (Biozyme); 4 uM proADAM-10 (Biozyme); 5% pro-domain buffer (Biozyme); 0.2% DMSO. The 12Z cell line was generously provided by Anna Starzinski-Powitz (University of Frankfurt) by way of Steve Palmer (EMD Serono). Cells were cultured in DMEM/F12 supplemented with 100 U/ml penicillin, 100 mg/ml streptomycin (Invitrogen; Carlsbad, CA), along with 10% fetal bovine serum (Atlanta Biologicals; Atlanta, GA) at 37°C, 5% CO2. ELISA kits were obtained from R&D Systems (Minneapolis, MN). For supernatant experiments, cells were seeded overnight at 80% confluency and treated the following day with 50ng/ml TNF-a (Peprotech; Rocky Hill, NJ). 24hrs after treatment, the cells were trypsinized and analyzed for cell count and viability using ViCell instrumentation after supernatant collection (Beckman Coulter; Brea, CA). No significant difference was observed in cell count or viability across the different treatment conditions.

B.3 Computational analysis

Droplet image sequences were interpreted in a program written in Matlab (Mathworks; Natick, MA) using a modified spot finding algorithm (Santella et al., 2010). Tracked droplets were filtered for appropriate diameter and fluorescence continuity over time. The previously described PrAMA algorithm best suits mixtures with

some prior knowledge of enzyme composition, allowing the inference procedure to be exactly- or over-determined mathematically (Lawson et al., 1995). This method was used to analyze purified enzyme mixtures (Fig. 3-2D-E, Fig. B-3-B-C) and conditioned media samples (Fig. 3-2F-I). For clinical application, we assumed no prior knowledge and applied PrAMA to the under-determined case where possible proteases outnumber the quantity of employed substrates. Inference accuracy using the Lawson-Hanson algorithm for non-negative least squares (NNLS) was validated on data from roughly 100 different mixtures of recombinant enzymes to support this slight mathematical adjustment (Fig. B.2B) (Miller et al., 2011). Results show the adjustment to be effective, consistent with previous work using the same NNLS algorithm to discern specific protease activities (Gosalia et al., 2006). For PLSDA, variables were mean-centered and variance-normalized across the set of samples. Forward-variable selection heuristically identified minimal combinations of variables to describe disease status. The algorithm iteratively added input variables to the PLSDA model if they improved model fit as determined by leave-one-out cross validation AUROC. For inclusion, input variables were required to exhibit PLSDA loadings of greater magnitude than their loading standard error. Standard errors for scores, loadings, and variable importance in the projection (VIP) statistics were calculated by jack-knifing (Efron et al., 1983). Prediction variance was computed according to Hoskuldsson (Hoskuldsson 1988). All results demonstrated significance ($p<0.05$) as determined by the permutation test. Briefly, we randomly shuffled disease status relative to the model inputs, and ran the variable-selection procedure for each of 1000 random permutations. Shuffled data yielded a lower AUROC value than that achieved by the actual data-set more than 95% of the time.

B.4 Clinical subjects and procedures

All laparoscopy patients provided informed consent in accordance with a protocol approved by the Partners Human Research Committee and the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects.

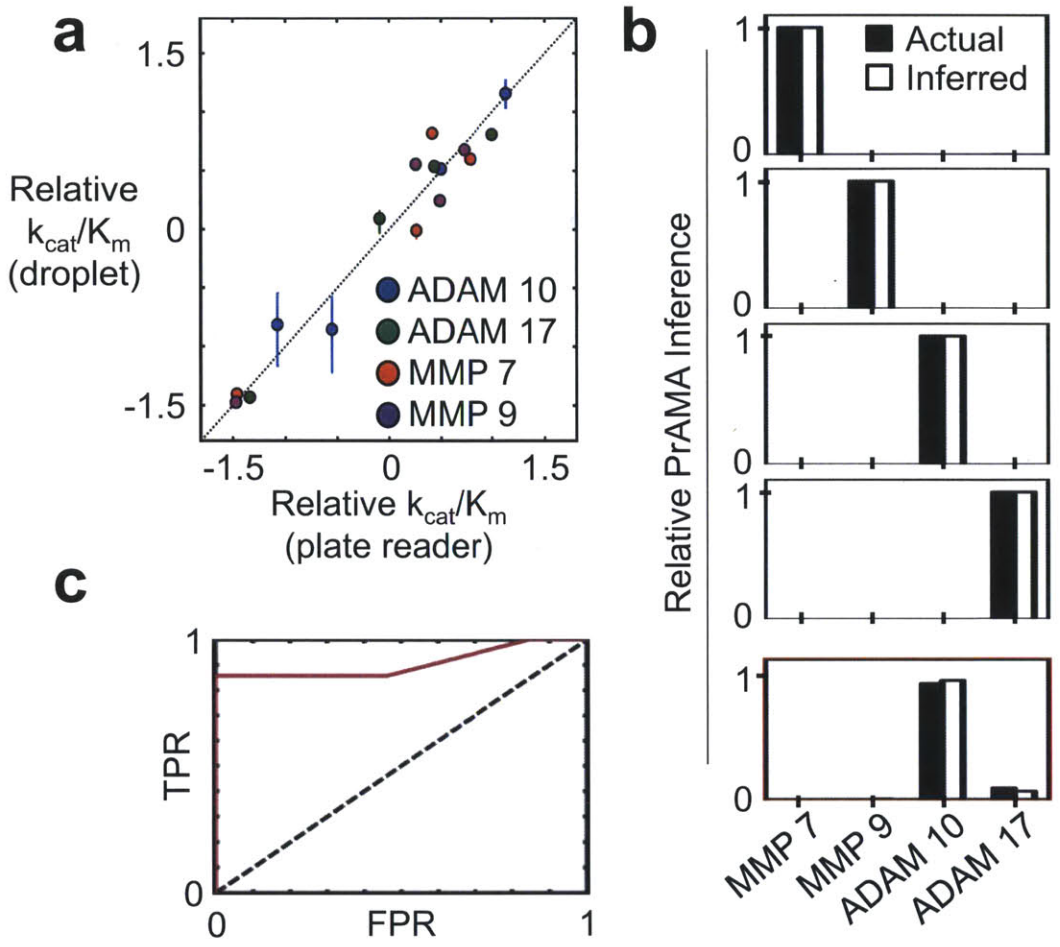

Figure B-1: **Validation of microfluidic PrAMA using plate-reader.** (a) Catalytic efficiencies obtained with the droplet platform (Fig. 3-2c) were compared to previously published measurements obtained in a traditional microtiter-plate format (R^2 >0.95). Rates were log-transformed, mean-centered, and variance normalized for each of the four enzymes. One of the sixteen data points (MMP-7 cleaving S1) was excluded as an outlier, lying >2 standard deviations from the average comparison error. (b) Top four boxes: PrAMA was used to infer the presence of single enzymes based on the observed cleavage signatures from the single-enzyme mixtures (Fig. 3-2c). Synthetic replicates of the data in Fig. 3-2c were simulated according to measured experimental inaccuracy, and PrAMA was used to infer specific enzymes based on these replicates. PrAMA inference parameters controlling sensitivity and specificity were optimized with this training set of data and applied to a mixture of two recombinant enzymes (bottom box, replicated from Fig. 3-2f). (c) Receiver-operator characteristic curve summarizing the overall accuracy of PrAMA across the five purified enzyme mixtures (AUROC = 0.9).

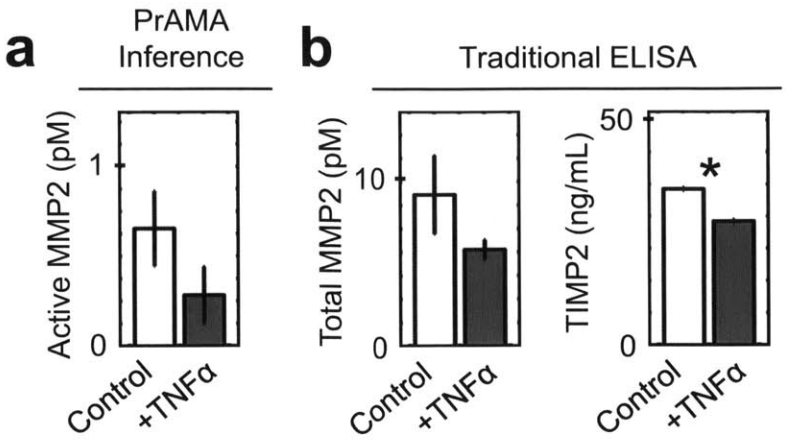

Figure B-2: **MMP-2 activity in 12Z conditioned media.** (a) PrAMA results for MMP-2 activity in response to TNF-a treatment (see Fig. 3-2g). (b) MMP-2 and TIMP-2 concentrations in the same conditioned media analyzed in (a) and Fig. 3-2g-i.

Enrollment was limited to pre-menopausal women with regular cycles (26-32 days) and excluded individuals who had received hormonal treatment within three months of surgery. Moderate/Severe (Stage III/IV) endometriosis was diagnosed according to the revised criteria of the American Society for Reproductive Medicine (Canis et al., 1997). Peritoneal fluid was aspirated from the rectouterine pouch immediately following trocar insertion and prior to peritoneal lavage or surgical manipulation. Upon collection, specimens were immediately placed on wet ice and clarified within 15 min. by centrifugation for 10 min. at 1000 rcf. Clarified aspirates were transported to the laboratory on ice, aliquoted, and stored at -80^{o}C until analysis. Two patient samples contained sufficient volume to serve as 'master-controls' that were also analyzed in a traditional microtiter-plate format. Enzymatic controls were performed with the 'master-controls' in the microtiter-plate to derive absolute cleavage rates from the microfluidic device measurements, through calculating a linear scaling factor for each of the nine reaction conditions (Miller et al., 2011). Variance in the scaling factor, across the two 'master-control' patient samples, described consistency between the two assay formats. For observed cleavage rates at least 50% above the background, median standard error in the scaling factor was calculated at 11%.

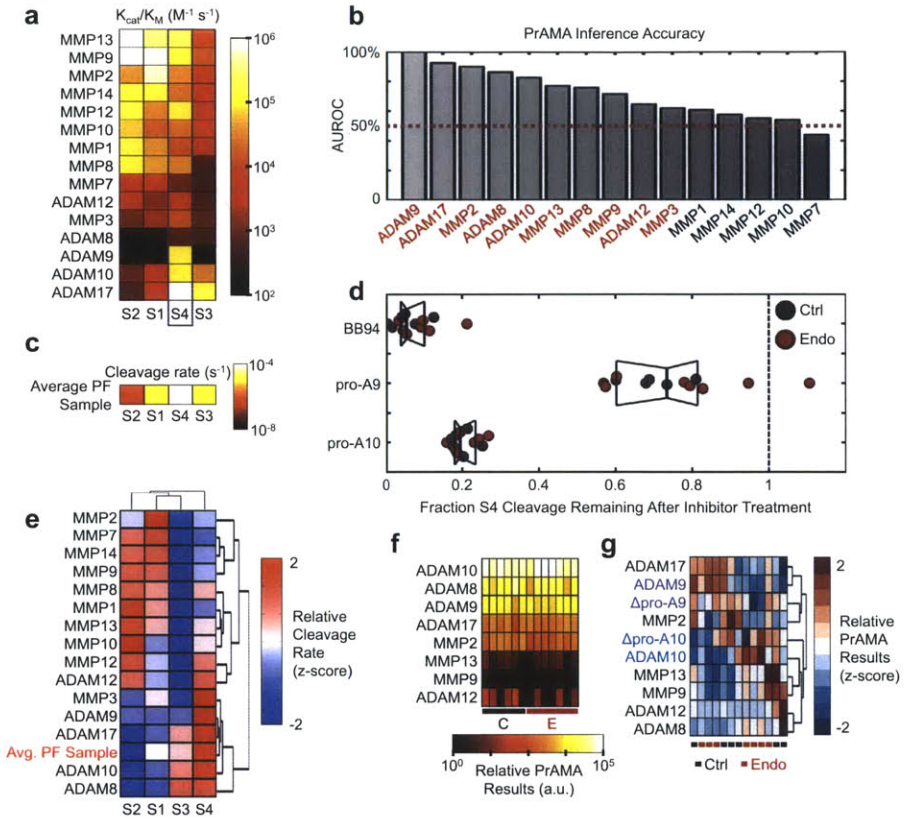

Figure B-3: **PrAMA inference accuracy and results.** (a) Catalytic efficiencies have been previously reported for substrates S1–S4 across a panel of recombinant enzymes, and are shown here in a clustergram (modified from Miller et al., 2011). (b) Overall PrAMA inference accuracy using the Lawson-Hanson algorithm and the parameter matrix shown in a. Cleavage signatures from roughly 100 different purified enzyme mixtures (reported previously) were applied to the PrAMA algorithm for under-determined inference. Red enzyme labels correspond to statistically significant inference performance ($p<0.05$, permutation test; Miller et al., 2011). (c) Average cleavage rate observed across all PF samples. (d) Effects of inhibitor treatment on observed S4 cleavage in PF samples. (e) Average PF sample cleavage rates were combined with purified enzyme rates, mean-centered and variance normalized by row, and hierarchically bi-clustered. PF sample cleavage patterns cluster closely with ADAM10 and ADAM17, both of which have high S4 cleavage efficiency. (f) Heat-map summarizing PF sample PrAMA inference results of data divided by the mean cleavage rate for each patient. (g) Hierarchical bi-clustering result corresponding to f, with data mean-centered and variance-normalized by row. ΔproADAM-9 denotes 1-[S4+pA9]/[S4+Buff-9], and ΔproADAM-10 denotes the analogous expression.

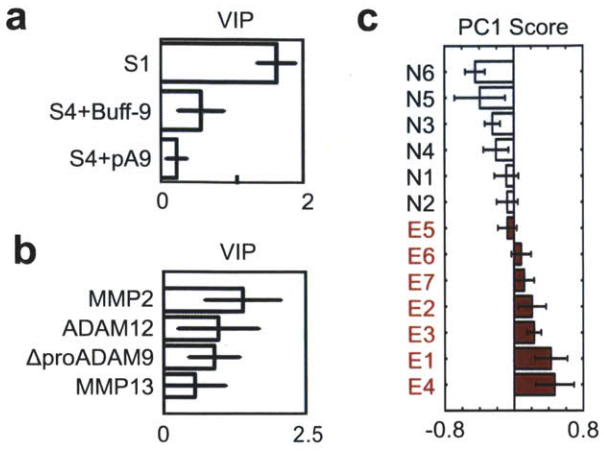

Figure B-4: **PLSDA and PrAMA modeling results.** (a) VIP scores for the PLSDA model describing disease status as a function of raw cleavage measurements (corresponding to Fig. 3b-c). (b) VIP scores for the PLSDA model describing disease status as a function of PrAMA results. (c) Scores plot for PLSDA results using PrAMA results, corresponding to b. For all plots, error bars represent standard error determined by jack-knife.

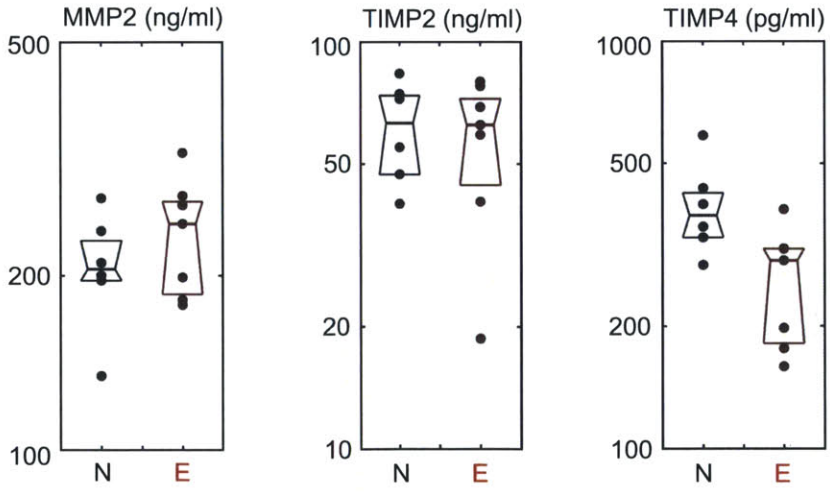

Figure B-5: **MMP/TIMP concentrations in PF samples.** MMP-2, TIMP-2, and TIMP-4 levels were measured by ELISA.

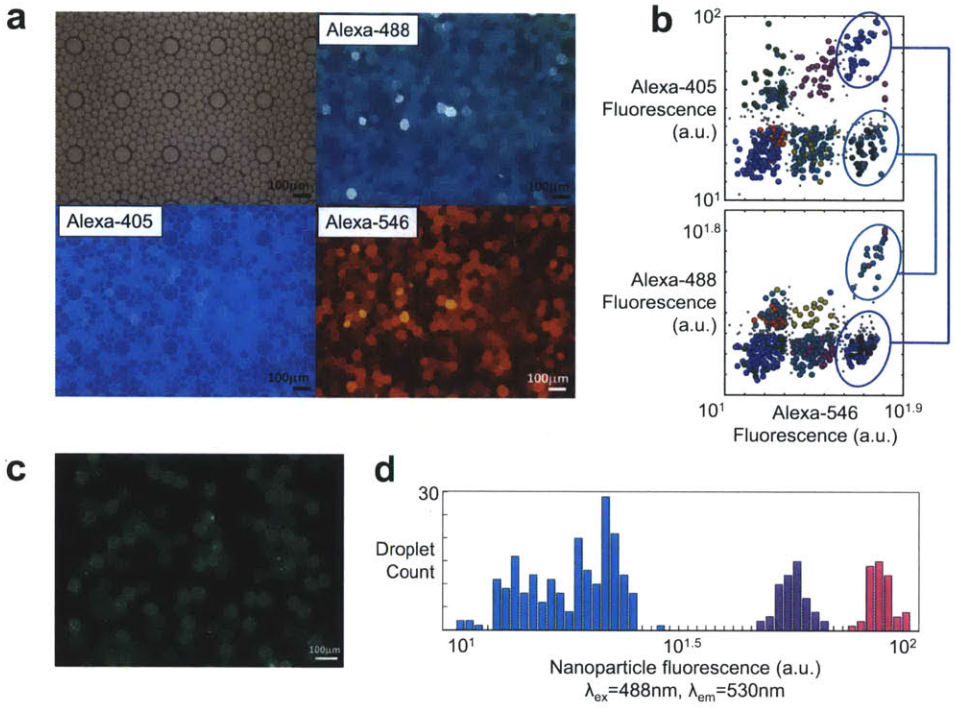

Figure B-6: **Additional droplet-library formulations.** (a) Formulation of a droplet-library developed with mixtures of three indicator dyes (Alexa-488, Alexa-405, and Alexa-546), shown by fluorescence microscopy. Three different concentrations of Alexa-488 and Alexa-546 (0ug/mL, 1ug/mL, 2ug/mL) and two different concentrations of Alexa-405 (0ug/mL, 1ug/mL) were prepared. Eleven different combinations of the three dyes at these concentrations are presented here. (b) Droplets were computationally identified and classified according to their prescribed combinations of indicator dyes. Droplet classifications are shown by color, as two different projections of droplet fluorescence from the three-dimensional space of fluorescence from the three indicator dyes. The cyan and blue circles indicate how droplets can be distinguished from one another by examining all three fluorescence values. Light gray dots remain unclassified and indicate droplets that fall outside of the gates for the eleven different possible dye combinations. (c) Formulation of a nanoparticle-labeled droplet library. Nanoparticles 100nm in diameter (λex=495nm, λem=530nm) were loaded into droplets at three different concentrations (0ug/mL, 2.5ug/mL, 5ug/mL) to encode a three-component droplet library. (d) Nanoparticle-labeled droplets exhibit distinct fluorescence values, shown as a histogram corresponding to c.

B.5 Discussion of clinical endometriosis findings.

In demonstrating our microfluidic platform advantages, results from the multiplexed assay provide a new perspective into MMP/TIMP regulation in endometriosis. The role of MMP activity in disease progression remains poorly understood. Several previous studies have found concentrations of MMPs, including MMP-2, to increase with endometriosis (Becker et al., 2010; Huang et al., 2004; Chung et al., 2002), while others have reported a slight decrease in MMP-2 levels in patients with severe disease (Protopapas et al., 2010). Furthermore, evidence of increased protease activity in the context of multiple interacting MMPs and TIMPs has proved even less consistent. Although increased MMPs suggest dysregulated endometriotic invasion, concomitantly increased TIMPs may indicate an overall decrease in protease activity that fits with the frequent observation of isolated endometriotic cysts that do not invade the surrounding tissue (Gilabert-Estelles et al., 2003). In this work, multiplexed protease activity measurements from endometriosis patients support the latter observation: although the concentration of MMP-2 may slightly increase with disease (B-5), the multivariate analysis reveals a much more significant and consistent decrease in relative MMP-2 activity (Fig. 3-3E). We additionally observed a decrease in relative ADAM-9 activity (Fig. 3-3D). Although relatively unexplored in the context of endometriosis, ADAM-9 has been implicated in hormonally-regulated biological processes including blastocyst implantation of the rabbit endometrium (Olson et al., 1998) and breast cancer progression, where for instance it has been shown to positively correlate with tamoxifen response (Sieuwerts et al., 2005). Standard ELISA measurements indeed confirmed dysregulated metalloproteinase/TIMP balance in the endometriosis PF samples, evidenced by a decrease in TIMP-4 with disease (Fig. 3-3F). Several previous studies have found clinical evidence of increased TIMP-4 with estrogen therapy (Pilka et al., 2006), endometrial hyperplasia (Pilka et al., 2004), and ovarian cancer (Ripley et al., 2006), while TIMP-4 levels may actually decrease in malignant endometrial cancer (Pilka et al., 2004). In endometriotic 12Z cells, TIMP-4 increases in response to inhibition of prostaglandin signaling (Lee et al., 2011). In

this work, we present the novel finding that relative TIMP-4 concentrations decrease in PF samples with disease. The non-intuitive and concomitant decrease of both MMP-2 and ADAM-9 activities, in spite of reduced TIMP-4 concentration, suggests that endometriosis perturbs multiple, overlapping protease-inhibitor interactions in the peritoneal environment. This complexity highlights the challenges associated with inferring enzyme activities from concentration alone, and emphasizes the need for direct activity measurements.

B.6 Discussion of more specific substrates and other possible applications.

The presented microfluidic platform could be extended to various applications. Low reagent and sample consumption make the platform amenable to non-invasive diagnostic screening (Roy et al., 2011). Fluorogenic substrates exist for diverse enzymes including cathepsins, caspases, and kinases (Stains et al., 2012), and droplet-based microfluidics could similarly improve liquid-handling and sample-quantity requirements in these applications. Strategies such as phage display (Felber et al., 2005) and positional scanning of synthetic libraries (Schneider et al., 2009) have shown promise to enhance the selectivity of substrates for proteases and other enzymes. However, for many applications issues of specificity remain problematic, for example with inherently promiscuous enzymes such as trypsin, and could still benefit from multiplexed approaches similar to PrAMA. As a further application, inhibitor activity of a sample could be assayed using droplet-libraries with spiked-in recombinant enzymes, analogous to the method of reverse zymography for TIMP activity measurement. The modularity among the droplet based microfluidic components, biochemical reagents, and methods to determine multiple enzyme activities ultimately makes the platform highly customizable for a variety of applications.

Appendix C

Supporting Information for Chapter 4.

Protein	Trend	Evidence	Citations
Stimuli			
TNFα	Up/–	S/PF/A	Kyama 2006; Bedaiwy 2002; Overton 1996
EGF	Up/–	PF/A	De Leon 1986; Huang 1996; Simms 1991
TGFα	Up	A	Simms 1991
PDGFbb	–	S/PF	Kalu 2007; Overton 1996
HGF	Up	PF	Osuga 1999
IGF1	Up/–	S	Matalliotakis 2003; Steff 2004; Gurgan 1999
Signaling			
p-p38	Up	T/A	Yoshino 2004; Yoshino 2006; Zhou 2010
p-Erk1/2	Up	T	Murk 2008; Honda 2008
p-Jnk	Up	T/A	Uz 2011; Yousef 2009
HSP27	Up	T	Matsuzaki 2004; El-Ghobashy 2005
c-Jun	Up	T	Shazand 2004; Ohlsson 2009
p-Akt	Up	T	Cinar 2009; Laudanski 2009; Honda 2008
p-STAT3	Up	T	Itoh 2013
IRS-2	N/A	GP	Cayan 2010
Tyk2	N/A	GP	Peluso 2012
RTK/Ligand			
EGFR	Up/–	T/S	Uzan 2009; Matalliotakis 2003; Ejskjaer 2009
TNFR1	Up	S/PF	Koga 2000; Steff 2004
MET	Up	T	Khan 2003
HER-2	Up/–	T/S	Prefumo 2003; Philippoussis 2004; Uzan 2009
AREG	Up	T	Aghajanova 2011
HBEGF	Up	T	Aghajanova 2011
MMP/TIMP			
MMP-2	Up	T/S/PF	Chung 2002; Huang 2004
MMP-3	Up	A/T	Ramon 2005; Bruner 2002; Saito 1995
MMP-7	Up	A/GP	Shan 2006; Bruner 2002
MMP-9	Up	T/PF	Chung 2001; Collette 2006;
TIMP-1	Up/Down	T/S/PF	Sharpe 1998; Ramon 2005; Gottschal 2002
TIMP-2	Down	T	Chung 2002; Gottschal 2002
ADAMs			
ADAM-17	Up	T	Gottschal 2002
ADAM-10	Up	T	Finas 2008

S = serum, PF = perit. fl., T = tissue, GP = gen. polymorph., A = animal model

Table C.1: **Table of key proteins and their roles in endometriosis.** Nearly all growth factors, cytokines, RTKs, ligands, and proteases directly measured in this work have been implicated in endometriosis to some degree by previously reported clinical data. "Trend" denotes increase or decrease with disease. Table is not meant to be comprehensive, but rather to demonstrate clinical relevance of study design.

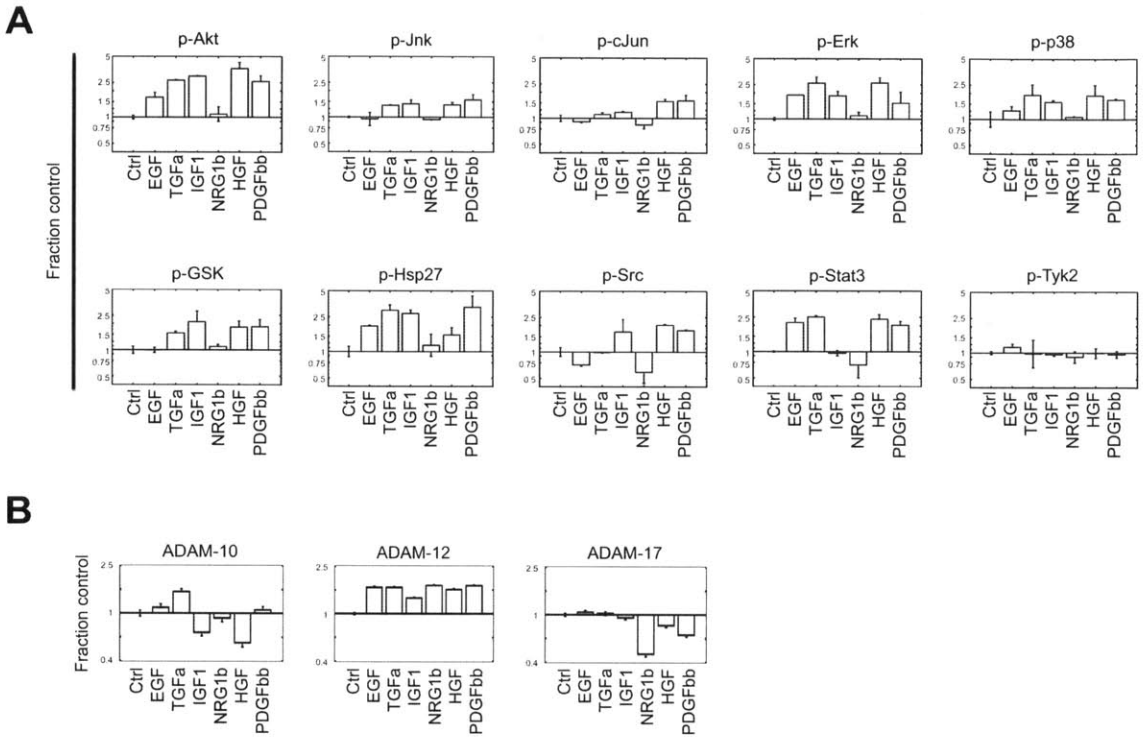

Figure C-1: **Phospho-signaling levels and proteolytic response to growth-factor treatment in 12Z.** Bar graphs correspond to data in Fig. 4-1A. (A) Phospho-signaling measurements, 5min post-stimulation, measured by bead-based ELISA and normalized to total protein content. (B) PrAMA results for ADAM-10, ADAM-12, and ADAM-17 measured over the first three hours post-stimulation. Error bars indicate standard error (n=2 experimental replicates for signaling data, n=4 replicates for PrAMA measurements).

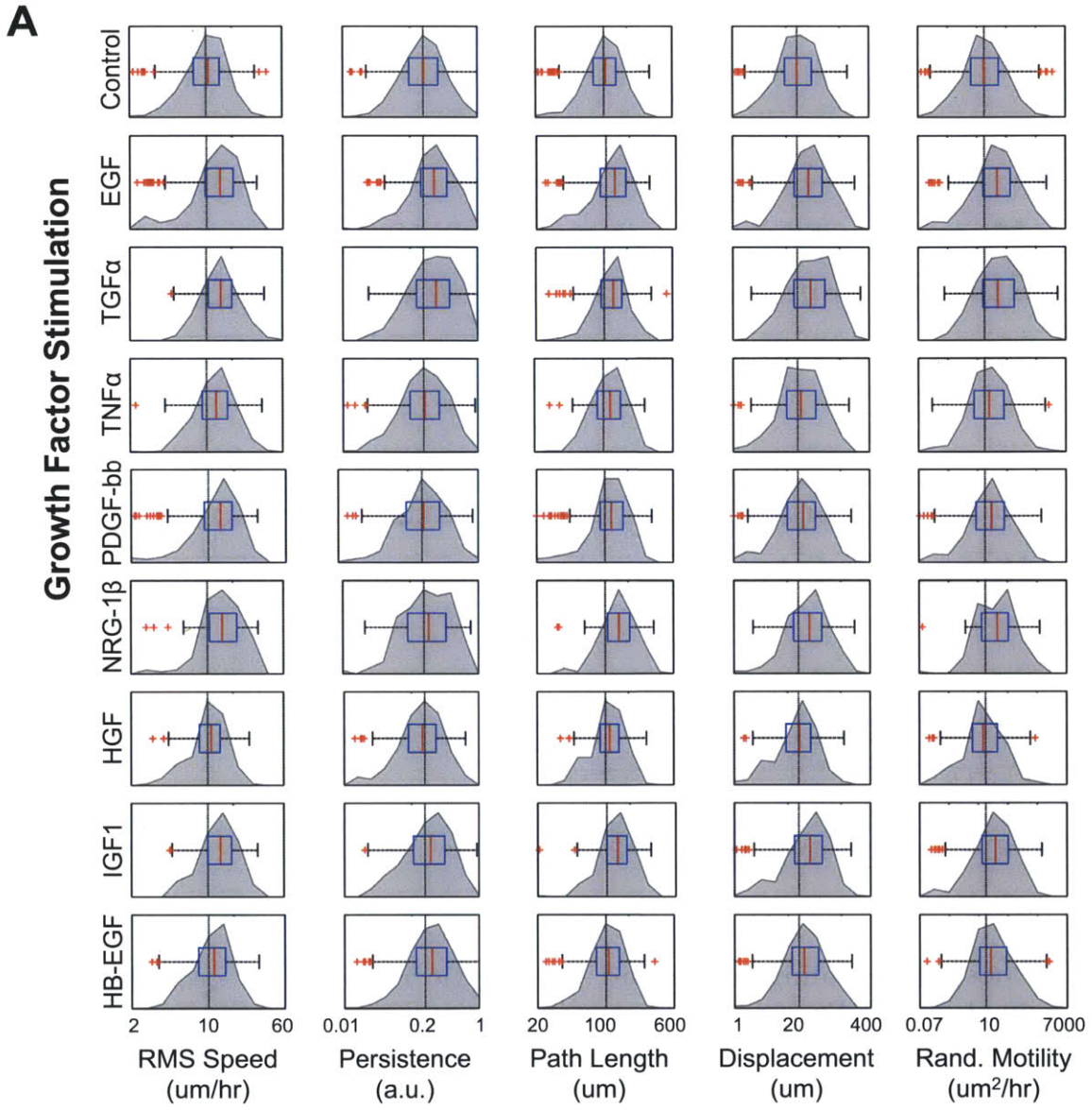

Figure C-2: **Single-cell motile response of 12Z to growth factor treatment.** Population distributions of single-cell motile response in 12Z. Various descriptive features of cellular motility were computed for individual cells based on single-cell tracking experiments using time-lapse confocal microscopy. Histograms and corresponding box-and-whisker plots show population distributions from single-cell measurements, pooled from n≥2 separate experiments, with ≥100 individual cells for each condition. A subset of these data are shown in Fig. 4-1B.

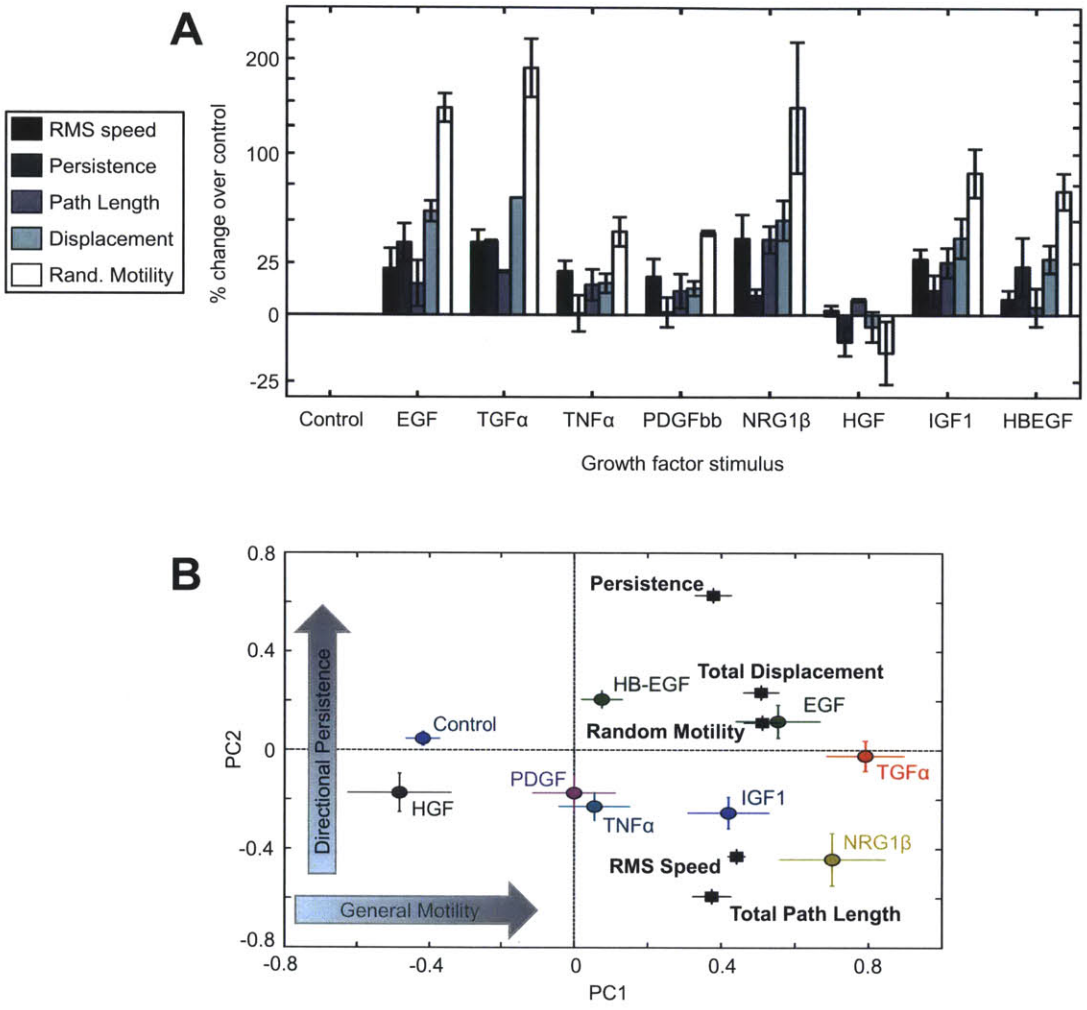

Figure C-3: **Multivariate analysis of 12Z motile response to growth factor/cytokine stimulation.** (A) Median population statistics were calculated from the single-cell data shown in Fig C-2, normalized to the control (+/- S.E.M. of experimental replicates, n≥2 separate populations, with ≥100 cells for each condition). (B) Principal component analysis of single-cell data decomposes observed motility features into two key axes of covariation, captured in the first and second principal components PC1 and PC2. Elliptical data-points correspond to scores of the different treatment conditions, and square data-points indicate principal component coefficients of the motility features (+/- S.E.M.). PC1 roughly describes general motility, while PC2 captures the directional persistence of motility.

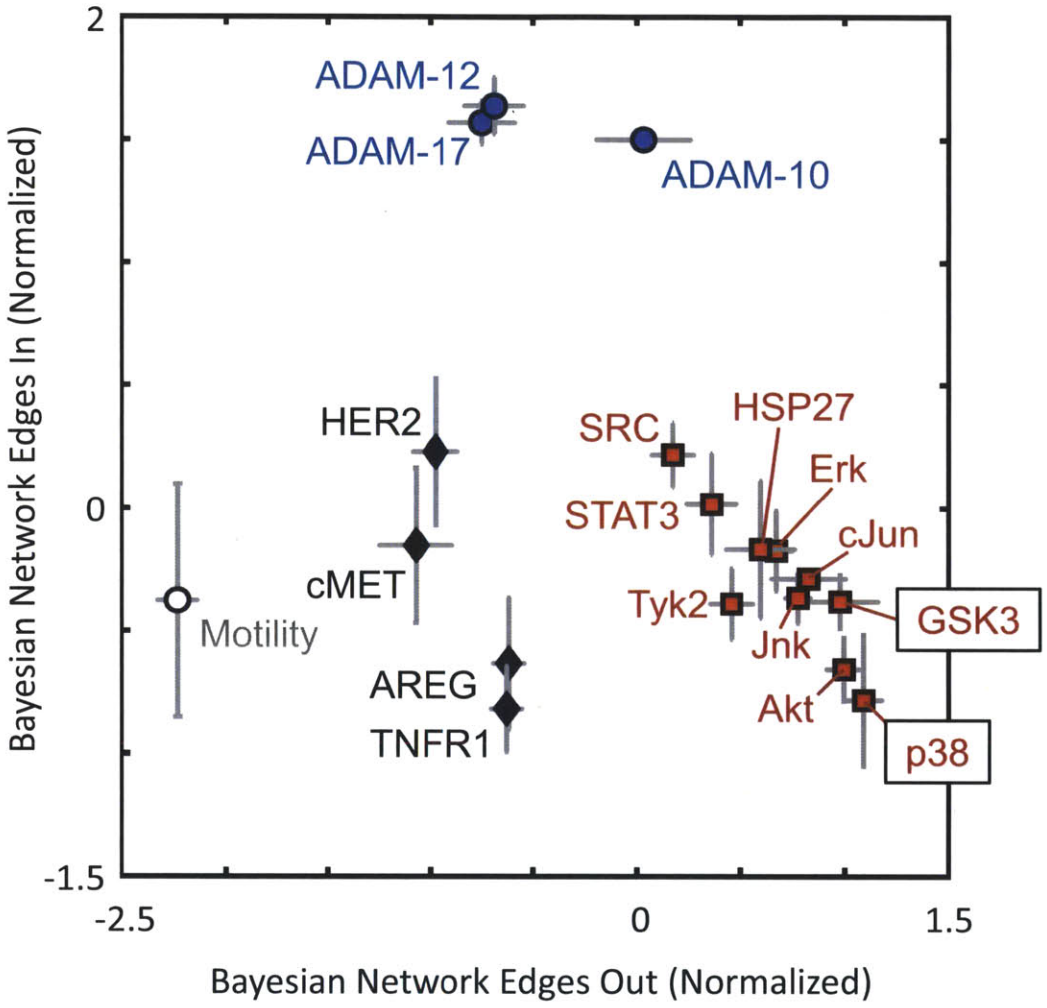

Figure C-4: **Bayesian network hierarchy analysis.** Bayesian network analysis inferred a hierarchical structure of signaling-mediated sheddase activity and motility, shown in Fig. 4-2C. In this figure, network nodes (comprising phospho-protein levels, protease activities, and supernatant analytes) are plotted based on the summed edge-weights leading into, and out from, each node. 'Upstream' nodes are in the lower right of the graph, and have more edges going out than coming into the node. Bayesian network analysis (Fig. 4-2C) represents an average of results across multiple Bayesian algorithm parameters, and therefore this plot shows average "edge-in" and "edge-out" values, +/- S.E.M., as they were calculated for each set of tested algorithm parameters. Values from individual Bayesian algorithm parameter sets were first mean-centered and variance normalized before averaging across multiple parameter sets.

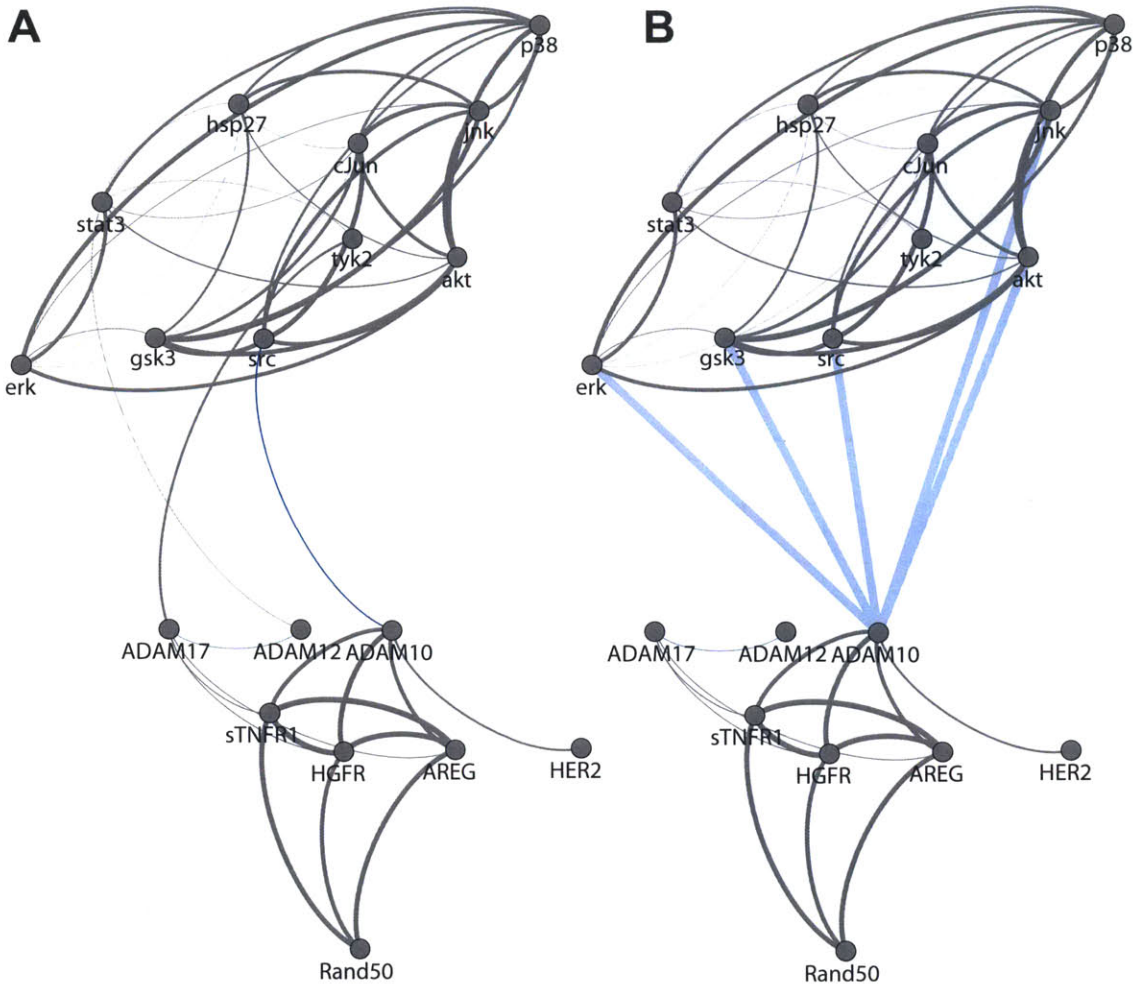

Figure C-5: **Bayesian network hierarchy analysis.** Data used in Bayesian network inference (Fig. 4-2C) were further investigated with correlation analysis and PLSR. (A) Pearson correlation was calculated between network nodes after mean-centering and variance normalization. Pairwise Pearson correlations above 0.4 are indicated in gray, while those below -0.4 are indicated in blue. Phospho-protein levels, protease activities, shedding measurements, and motility were considered as separate groups. Only edges within or between adjacent groups are shown here. (B) PLSR modeling describes ADAM activity as a function of phospho-protein levels. PLSR models were built using all possible combinations of five individual phospho-protein levels to describe each protease activity. The significantly enriched variables (shown by blue edges; $p<0.05$, enrichment score defined in Sec. C.1) were then included in a PLSR model to predict ADAM-10 activity ($Q^2 = 0.79$). No sufficiently predictive models of any size could be attained for ADAM-17 or ADAM-12 activity.

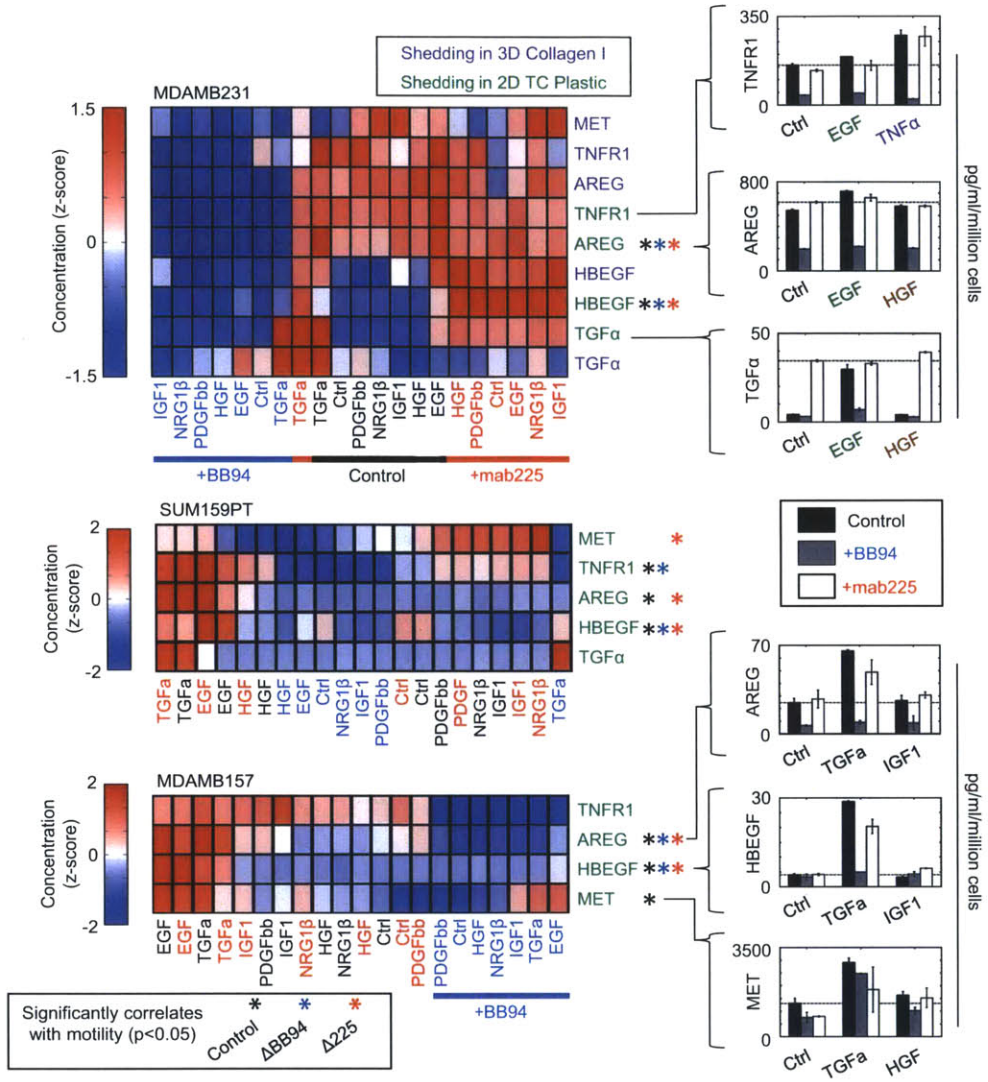

Figure C-6: **Ligand and RTK shedding in TNBC cell lines.** MDAMB231, SUM159PT, and MDAMB157 were stimulated with the same growth factor treatments as those performed in the 12Z studies, in the presence and absence of BB94 or mab225. For MDAMB231, some analytes were measured in 2D environments with cells growing on tissue-culture plastic, or in 3D environments with cells growing in suspension cultures of collagen I gel. In the heat-maps, measurements were first mean-centered and variance-normalized by row. We examined correlation between ectodomain shedding and cell motility using previously published data describing the 3D motile response of cells to the same growth factor cues (Meyer et al., 2012). Asterisks denote significant correlation, after FDR correction (Storey et al., 2002). Δ225 and ΔBB94 denote the change in analyte accumulation observed with inhibitor treatment, across the panel of growth factor conditions.

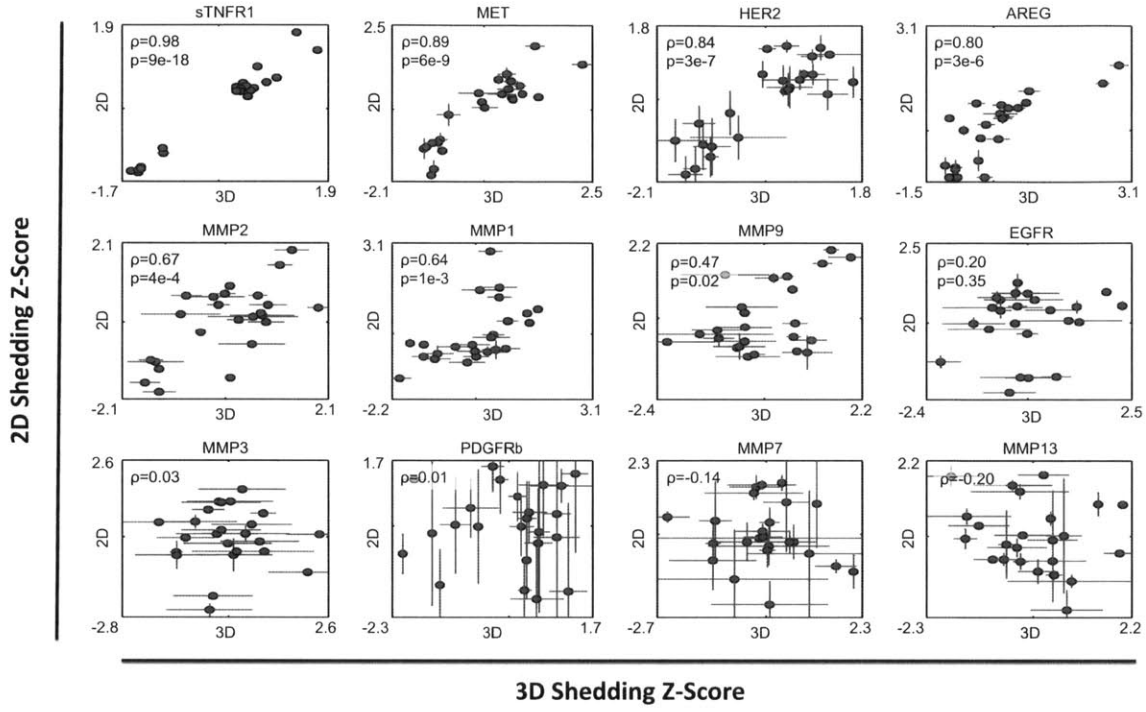

Figure C-7: **Shedding measurements in 2D vs. 3D cell environments.** MMP secretion and ectodomain shedding were measured in 12Z grown on either tissue culture plastic or in collagen I gels. Supernatant from both cell culture environments were collected 24hrs post-stimulation with the panel of growth factors used in Fig. 4-1, in the presence or absence of mab225 and BB94. Error-bars indicate standard error. Spearman's and Pearson's correlation coefficients were calculated for each analyte, and the greater of the two are reported as ρ. P-values based on correlation were calculated using either a Student's t distribution or an approximation of the exact permutation distributions, depending respectively on whether ρ denotes Pearson's or Spearman's correlation.

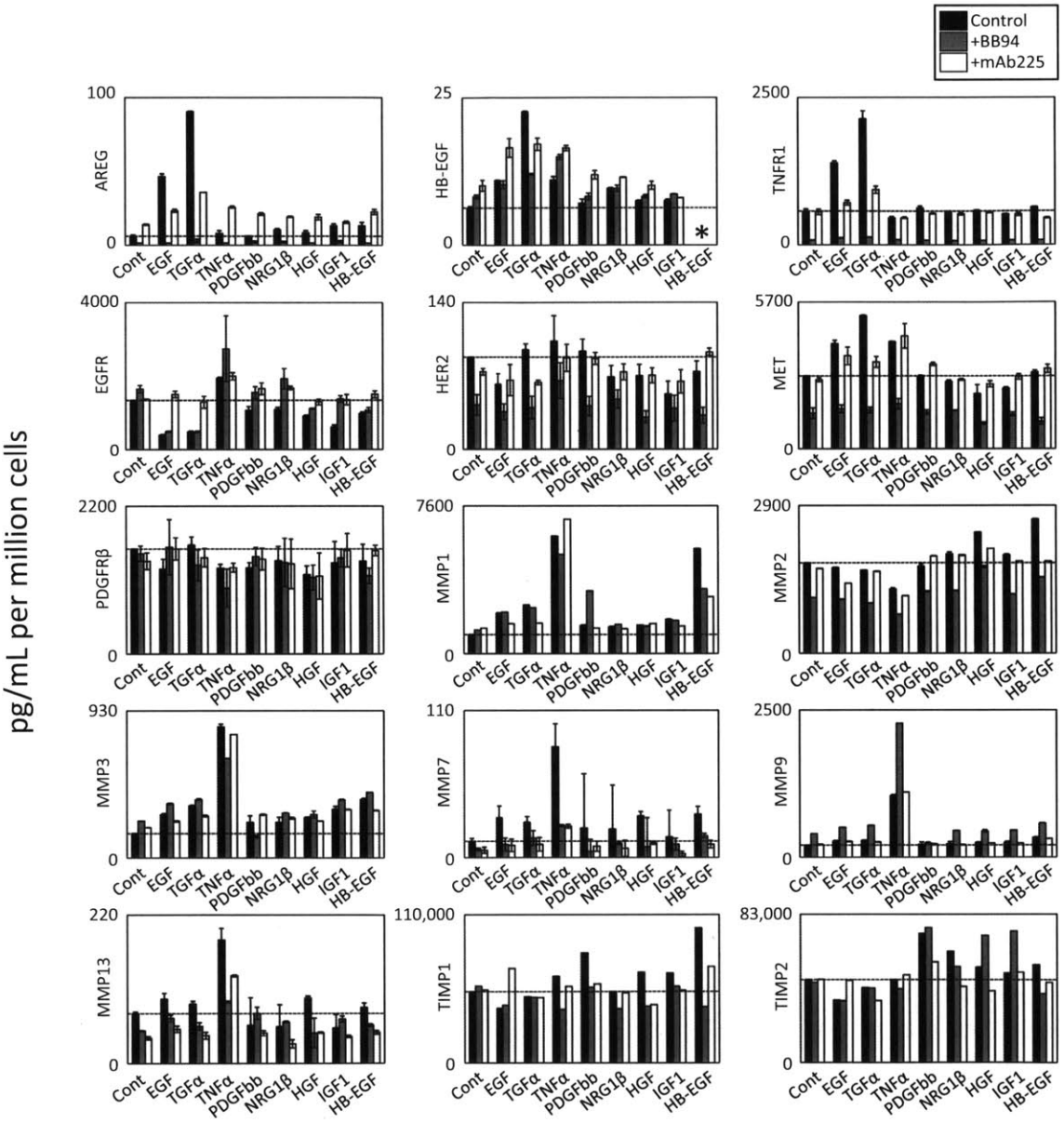

Figure C-8: **12Z supernatant proteomics with mab225 and BB94 treatment.** Supernatant analyte concentrations were measured 24hrs post-stimulation from 12Z grown on tissue culture plastic in the presence of either 10μM BB94 or 10μg/mL mAb225. Error bars indicate ± S.E.M., with n=2 experimental replicates. TIMPs have n=1 replicate. HB-EGF measurements in the presence of exogenous HB-EGF are excluded due to assay saturation.

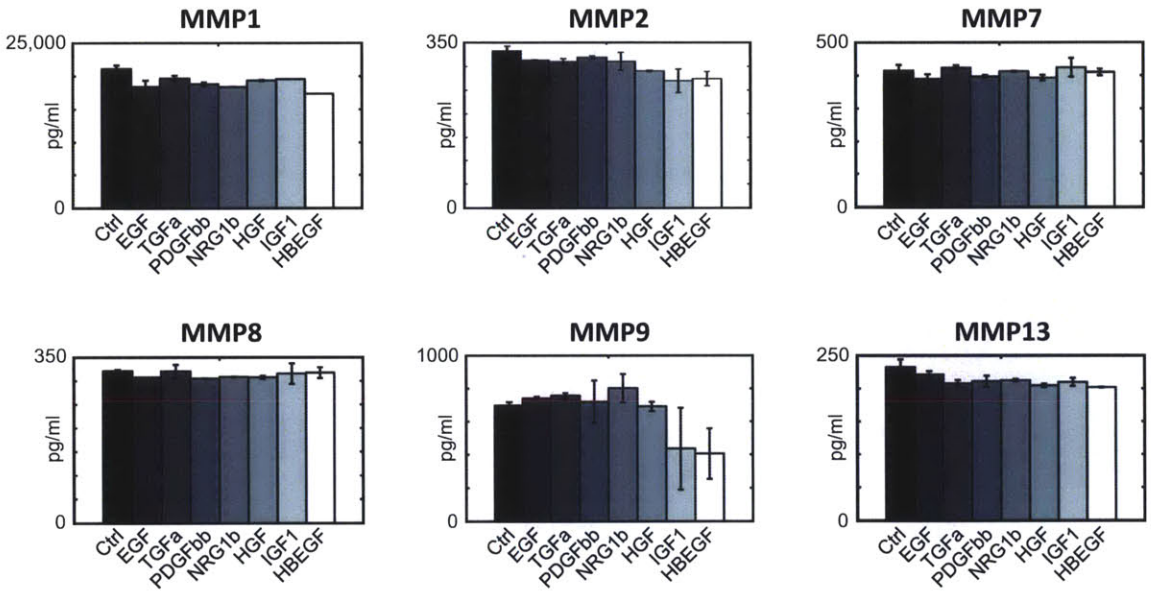

Figure C-9: **MMP secretion in MDAMB231.** MMP accumulation was measured in the supernatant of MDAMB231 growing in suspension cultures of collagen I gel. Supernatants were collected 24hrs after growth-factor stimulation. MMP secretion does not significantly correlate with motility in MDAMB231.

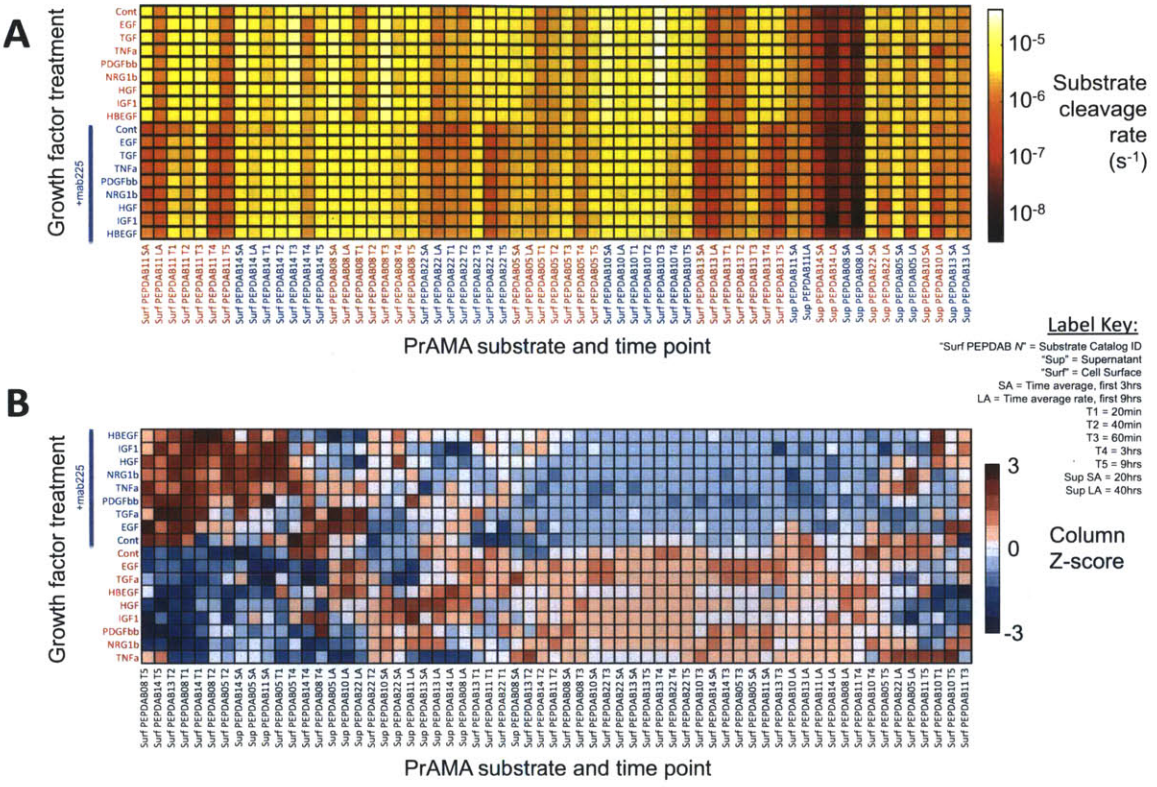

Figure C-10: **Complete 12Z FRET-substrate cleavage data.** PrAMA was used to measure the effects growth factors/cytokines and mab225 treatment on metalloproteinase activities. After a 15min pre-treatment with mab225, 12Z cells were co-treated with growth factor and PrAMA FRET-substrates. (A) Raw cleavage rates of seven PrAMA substrates at multiple time-points, on the cell surface and in the supernatant. For supernatant activity measurements, PrAMA substrates were added to supernatant that was removed from cells and clarified at 24hrs post-treatment. Fluorescence was measured at 20 and 40hrs after substrate addition. (B) Clustergram of results from *A*.

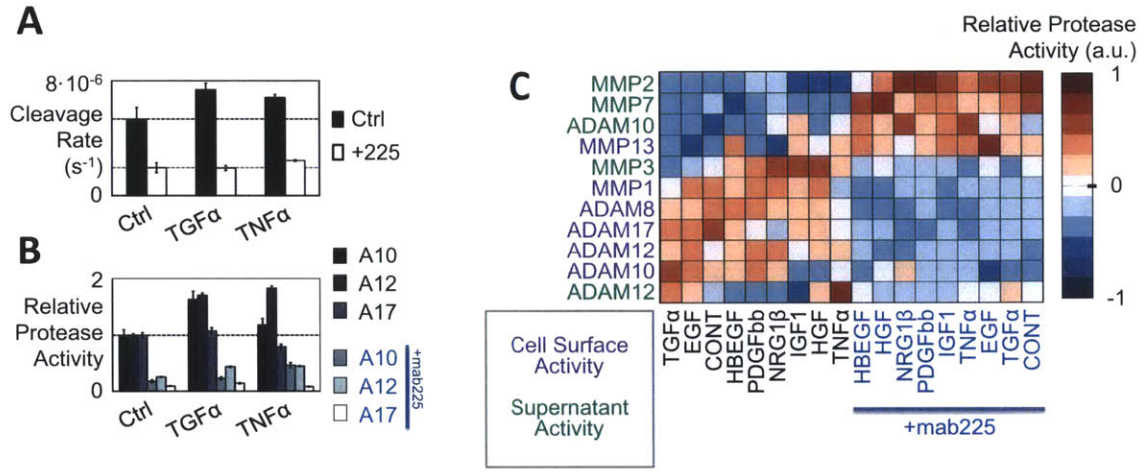

Figure C-11: **Summary of complete 12Z PrAMA measurements.** PrAMA inference results corresponding to Fig. C-10. PrAMA was used to measure the effects growth factors/cytokines and mab225 treatment on metalloproteinase activities. After a 15min pre-treatment with mab225, 12Z cells were co-treated with growth factor and PrAMA substrates. (A) Raw PrAMA substrate shedding of PEPDAB05, which is efficiently cleaved by ADAMs (Miller et al., 2011), and corresponds to "Surf PEPDAB05 SA" in Fig. C-10A. (B) PrAMA results were determined using the data from Fig. C-10A averaged over the first 3hrs post-stimulation ("SA" time-point), to directly infer metalloproteinase activity. (C) Clustergram of PrAMA results, including those presented in *B*. Cell-surface ADAM-10, ADAM-12, and ADAM-17 measurements were used in the initial network-inference analyses (Fig. 4-1 & 4-2). PrAMA algorithm parameters have been previously described (Chen et al., 2012).

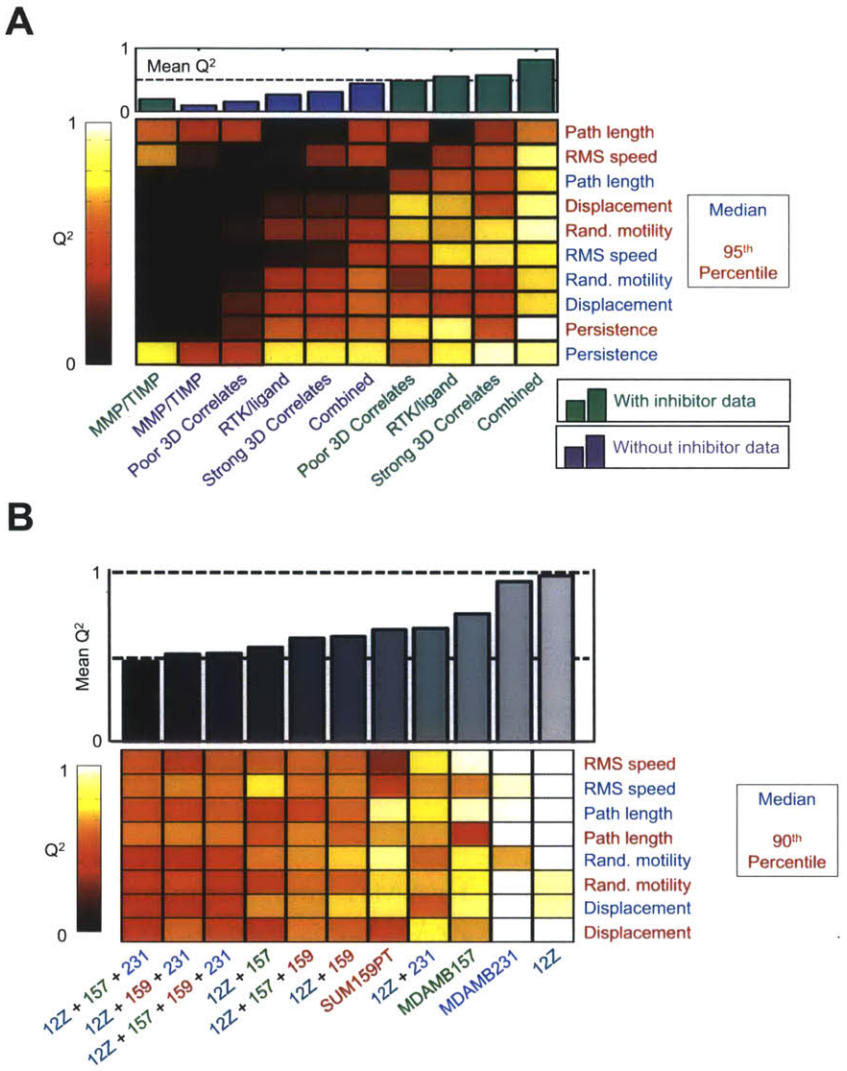

Figure C-12: **PLSR uses ectodomain shedding to predict motility.** (A) Q^2 accuracies are reported for PLSR models that use various subsets of 12Z supernatant measurements (columns) to describe various features of 12Z motility, for either the median or 95th percentile of single-cells in the population. The bar plot shows column averages corresponding to the heat-map. Growth factor and cytokine treatments tested in these models are described in Fig. C-2. (B) Q^2 accuracies are reported for PLSR models that describe shedding-motility relationships for individual and combinations of cell lines. Combined cell line models use the same statistical parameters to describe motility across multiple cell lines and growth factors, whereas the individual cell line models use PLSR parameters that are specific to the particular cell line being described. The bar plot shows column averages corresponding to the heat-map. Shedding measurements for the models comprise all overlapping sets of measurements reported in Fig. C-6 & C-8.

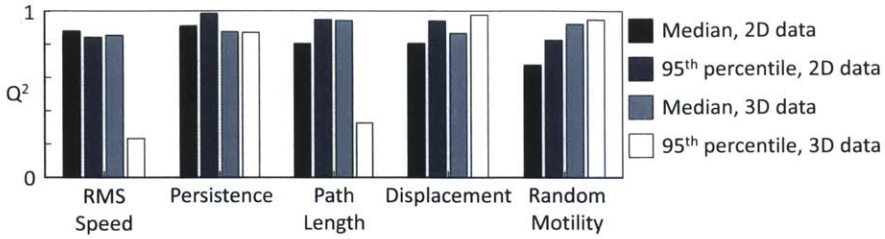

Figure C-13: **Ectodomain shedding measurements from both 2D and 3D cell cultures equally predict 3D motility.** PLSR models were generated to predict 3D 12Z motility using supernatant measurements collected from cells cultured on either tissue culture plastic (2D) or from cells suspended in collagen I gels (3D). According to Q^2 fitting accuracy, PLSR is capable of using both 2D and 3D supernatant measurements to accurately predict 3D cellular motility. Enhanced signal-to-noise in 2D shedding measurements may explain higher prediction accuracy for some cases. These results are concordant with observed correlation between supernatant measurements in 2D and 3D (Fig. C-7).

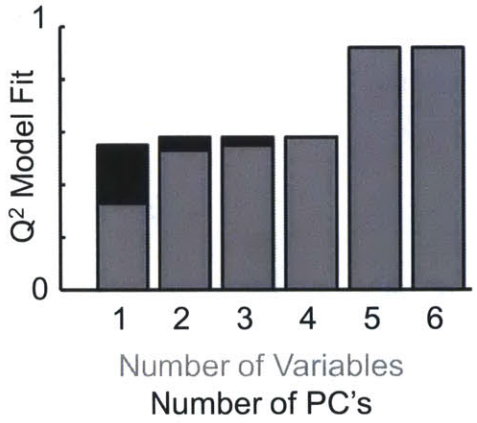

Figure C-14: **Accurate prediction of motility requires multiple principal components.** The integrated PLSR model that describes relationships between ectodomain shedding and motility in both 12Z and MDAMB231 (Fig. 4-4A) depends upon multiple shedding measurements and principal components. In this figure, limits are placed on the number of principal components (shown as black bars) or the number of variables (shown as gray bars, laid over the black bars) allowed to be incorporated into the model structure.

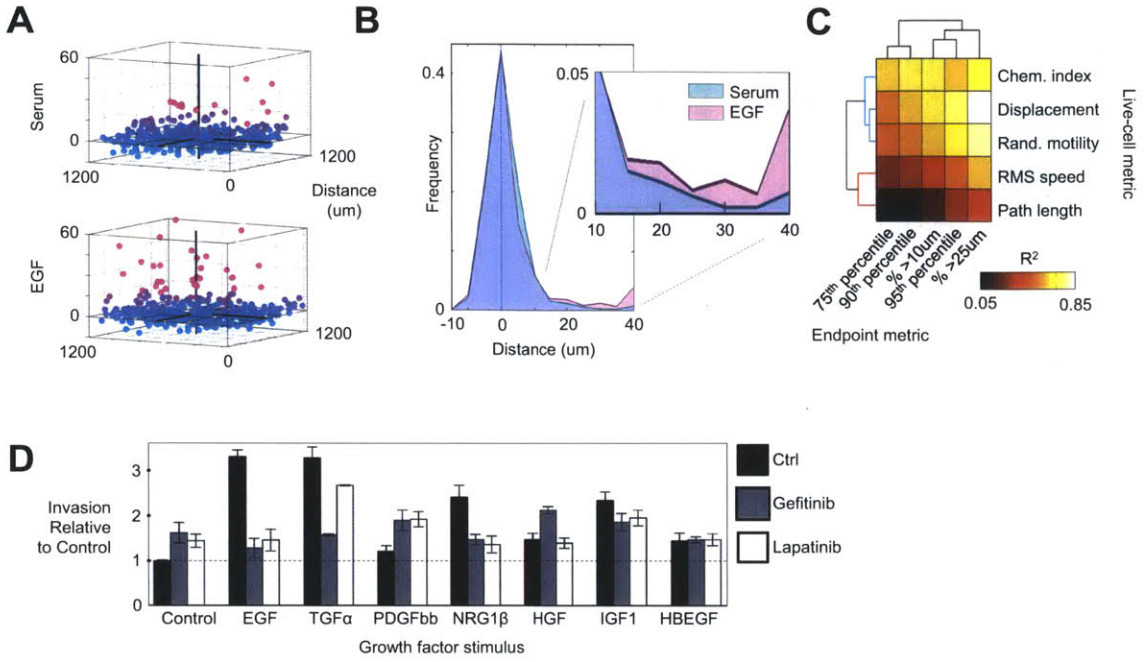

Figure C-15: **Endpoint invasion assay correlates with live-cell migration.** (A) Computationally inferred cell nuclei positions from the invasion assay, in the presence or absence of EGF. The x-y plane represents the inferred bottom of the 96-well plate. Nuclei with large z-coordinates are colored warmer. (B) Distribution of nuclei z-coordinates, corresponding to *A*. (C) Clustergram of correlations between the endpoint assay metrics and median live-cell motility data, compared across the panel of growth factors.

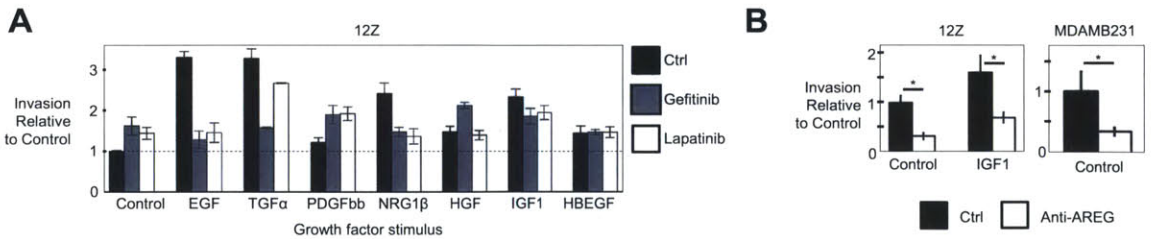

Figure C-16: **EGFR pathway inhibitors block invasion.** Endpoint invasion assays were performed following growth factor and inhibitor treatments. (A) 12Z were stimulated with growth factors following a pre-treatment with the EGFR kinase inhibitor, gefitinib, or the dual EGFR/HER-2 kinase inhibitor, lapatinib. Gefitinib results correspond to Fig. 4-5B. (B) 12Z were stimulated with IGF1 following a pre-treatment with an AREG decoy antibody. Basal invasion was also measured in MDAMB231, with decoy antibody treatment. Error bars indicate S.E.M. from two separate experiments. In *B*, asterisks indicate $p<0.05$ as determined by single-tailed Student's t-test.

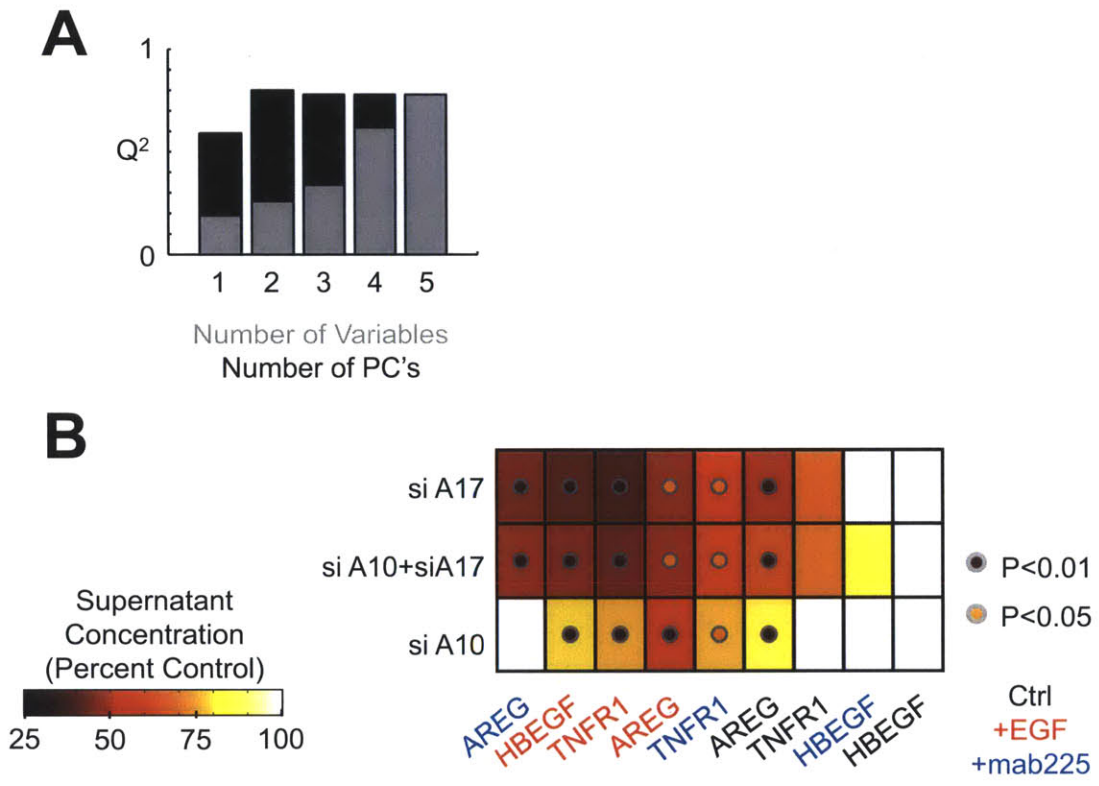

Figure C-17: **PLSR and siRNA suggest cross-talk between ADAM-10 and ADAM-17.** (A) PLSR was used to build models that describe AREG accumulation (shown in Fig. C-8) as a function of PrAMA-determined ADAM activities (shown in Fig. C-11). PLSR model accuracy improves with the inclusion of multiple protease activities and principal components, suggesting that AREG shedding is regulated by multiple proteases. Results from the model using three variables and two principal components (3 VAR, 2 PC) are shown in Fig. 4-6B. (B) siRNA knockdown of both ADAM-10 and ADAM-17 reduces shedding of multiple endogenous substrates. Results correspond to Fig. 4-6C.

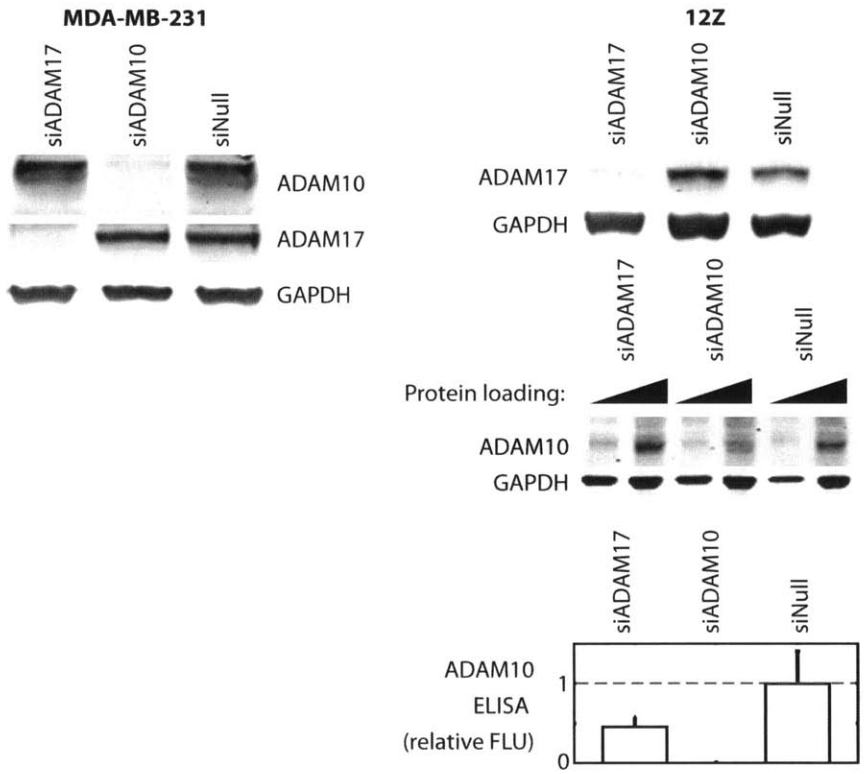

Figure C-18: **Validation of siRNA ADAM-10/ADAM-17 knockdown by western blot and ELISA.** Western blots show ADAM-10 and ADAM-17 knockdown in MDAMB231 (left side) and 12Z (right side) following siRNA treatment, at 48hrs post-transfection. For ADAM-10 measurement in 12Z, high western-blot background prompted the use of a bead-based sandwich ELISA for ADAM-10. Bead fluorescence was quantified and shown relative to control levels. For the ELISA, statistically significant decrease in signal was observed only upon siADAM-10 treatment ($p<0.05$, Student's t-test).

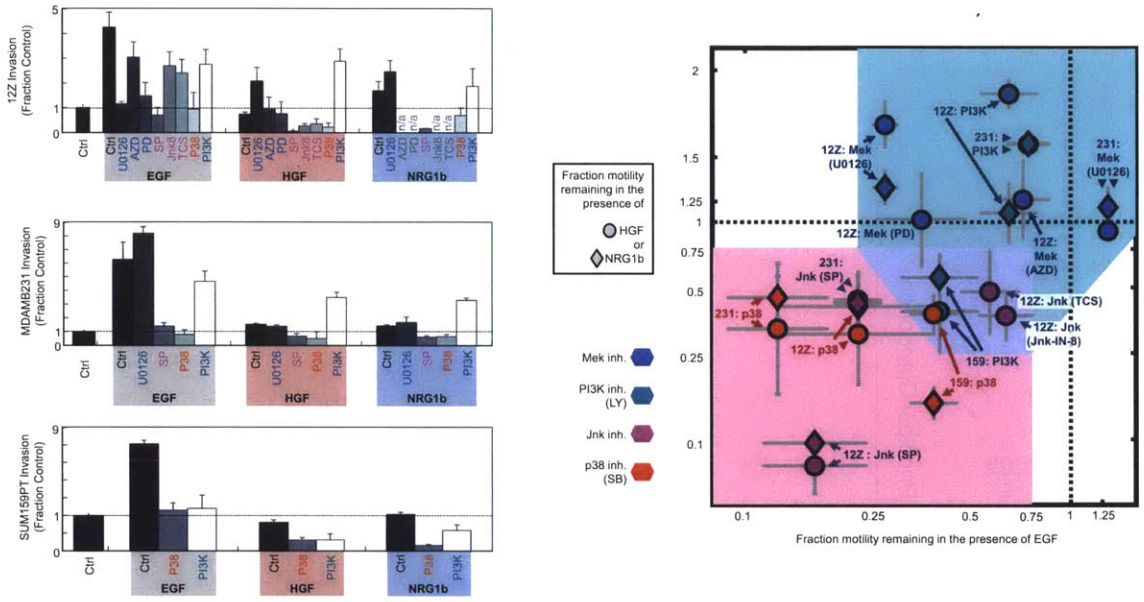

Figure C-19: **Cellular invasion response to growth factor and kinase inhibitor treatment.** The endpoint cellular invasion assay was performed with three cell lines and a variety of kinase inhibitors. Data corresponds to Fig. 4-7A-B. "n/a" indicates experimental conditions were not tested. Magenta lettering indicates Jnk inhibition, and blue lettering indicates Mek inhibition. The scatterplot at right corresponds to the bar plots at left, and shows the fraction motility remaining after inhibitor treatment, in the presence of EGF and either HGF or NRG1b.

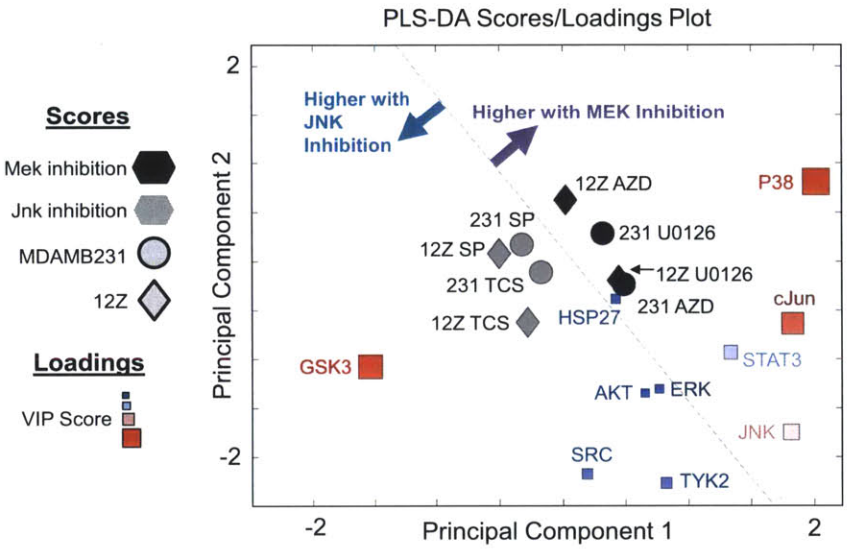

Figure C-20: **Fully labeled PLS-DA scores and loadings plot for comparing the signaling responses between cells treated with either Mek or Jnk inhibitors.** Corresponds to Fig. 4-8B, with full labels.

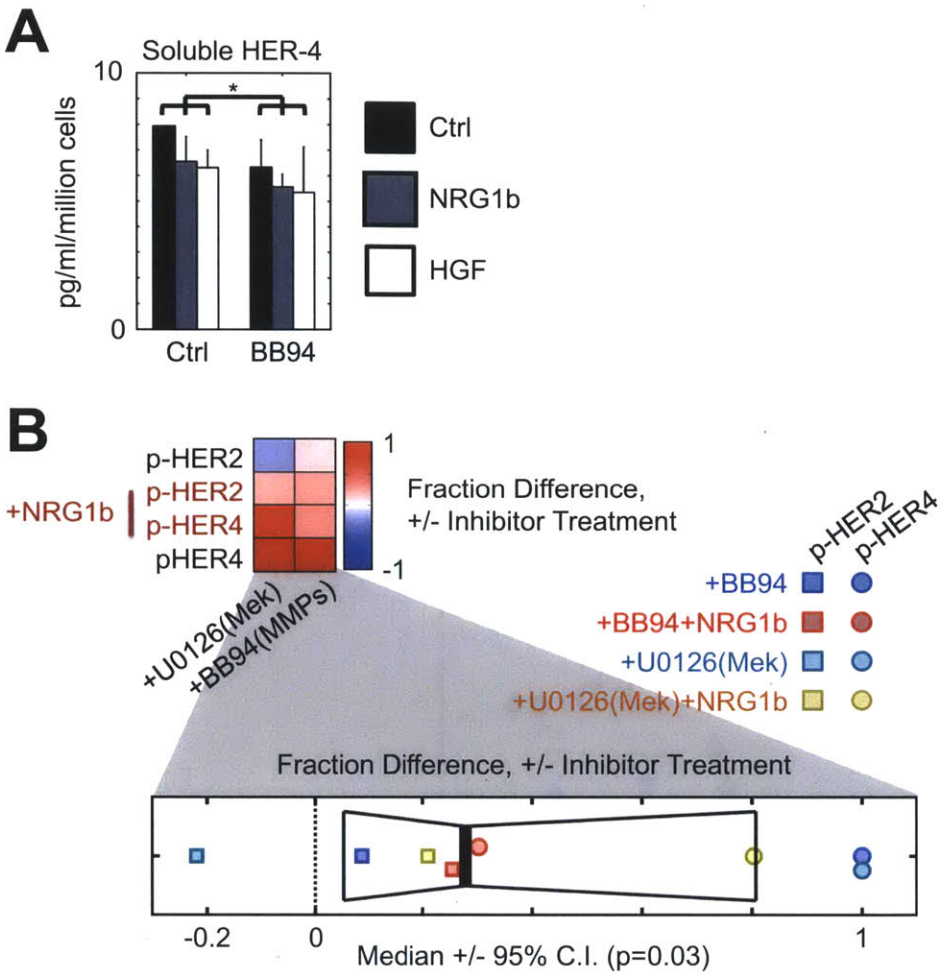

Figure C-21: **Kinase and protease inhibitors affect RTK shedding in 12Z and MDAMB231.** (A) BB94 treatment reduces supernatant HER-4 in 12Z, measured at 24hrs post-treatment ($p<0.05$, paired Student's t-test). (B) U0126 and BB94 increase p-HER-2 and p-HER-4 in MDAMB231, measured 5min post-stimulation with NRG1b, after a 1hr pre-treatment with inhibitor. Each data point represent the average of two experimental replicates. Across the eight data points shown, phosphorylation of full-length receptor exhibits a statistically significant increase of roughly 30% with U0126 or BB94 treatment ($p<0.05$, paired Student's t-test).

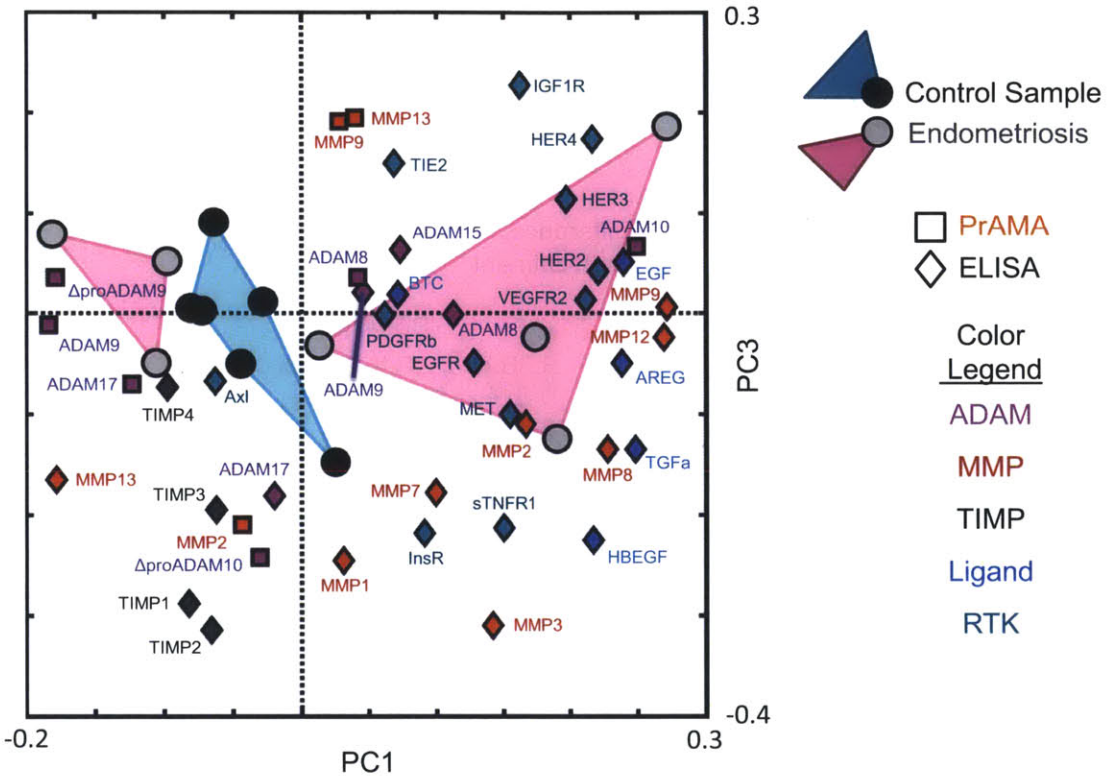

Figure C-22: **Fully labeled PCA scores/loadings plot of PF samples.** PCA decomposes the ELISA and PrAMA measurements of PF samples into key principal components of variation. Scores and loadings for PC1 and PC3 are shown here, identical to Fig. 4-10B, but with complete labeling.

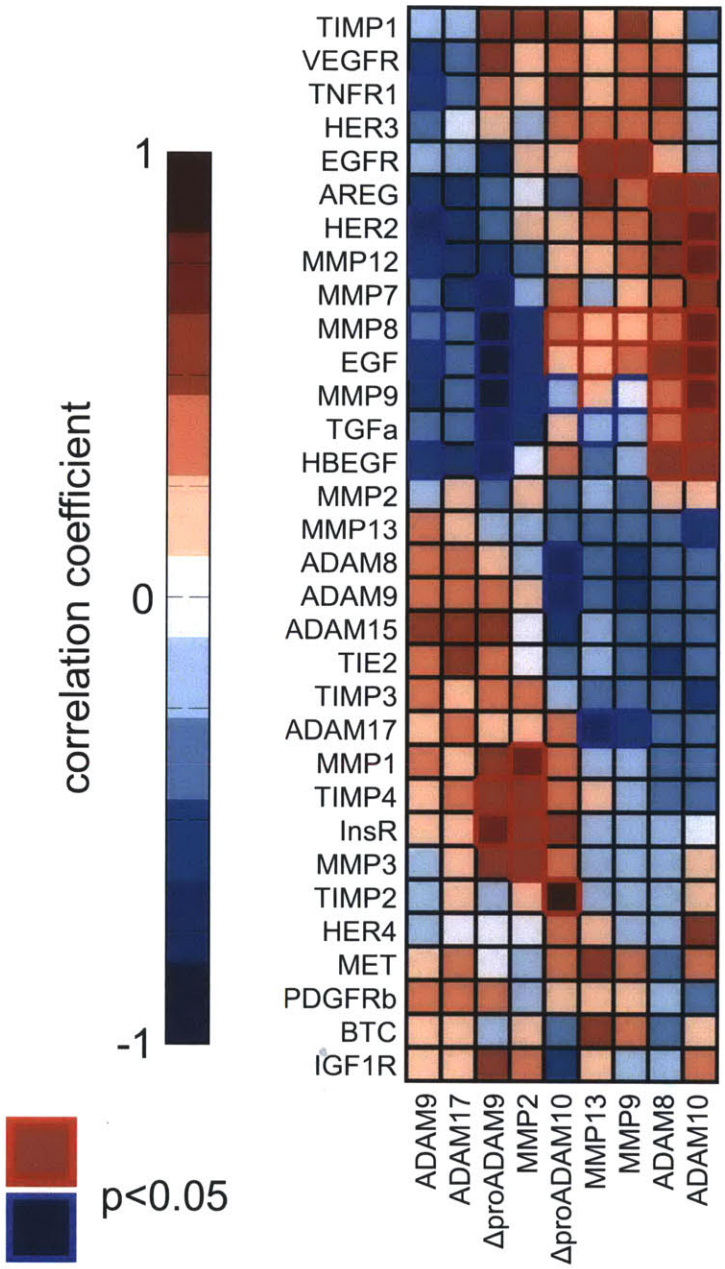

Figure C-23: **Correlation between PF protease activity and analyte concentrations.** Spearman correlation was calculated between protease activities (columns) and observed analyte concentrations (rows) across the set of thirteen PF samples. The t-test for significance was determined using an approximation of the exact permutation distributions of correlation, with Storey correction for a 0.05 FDR (Storey et al., 2002).

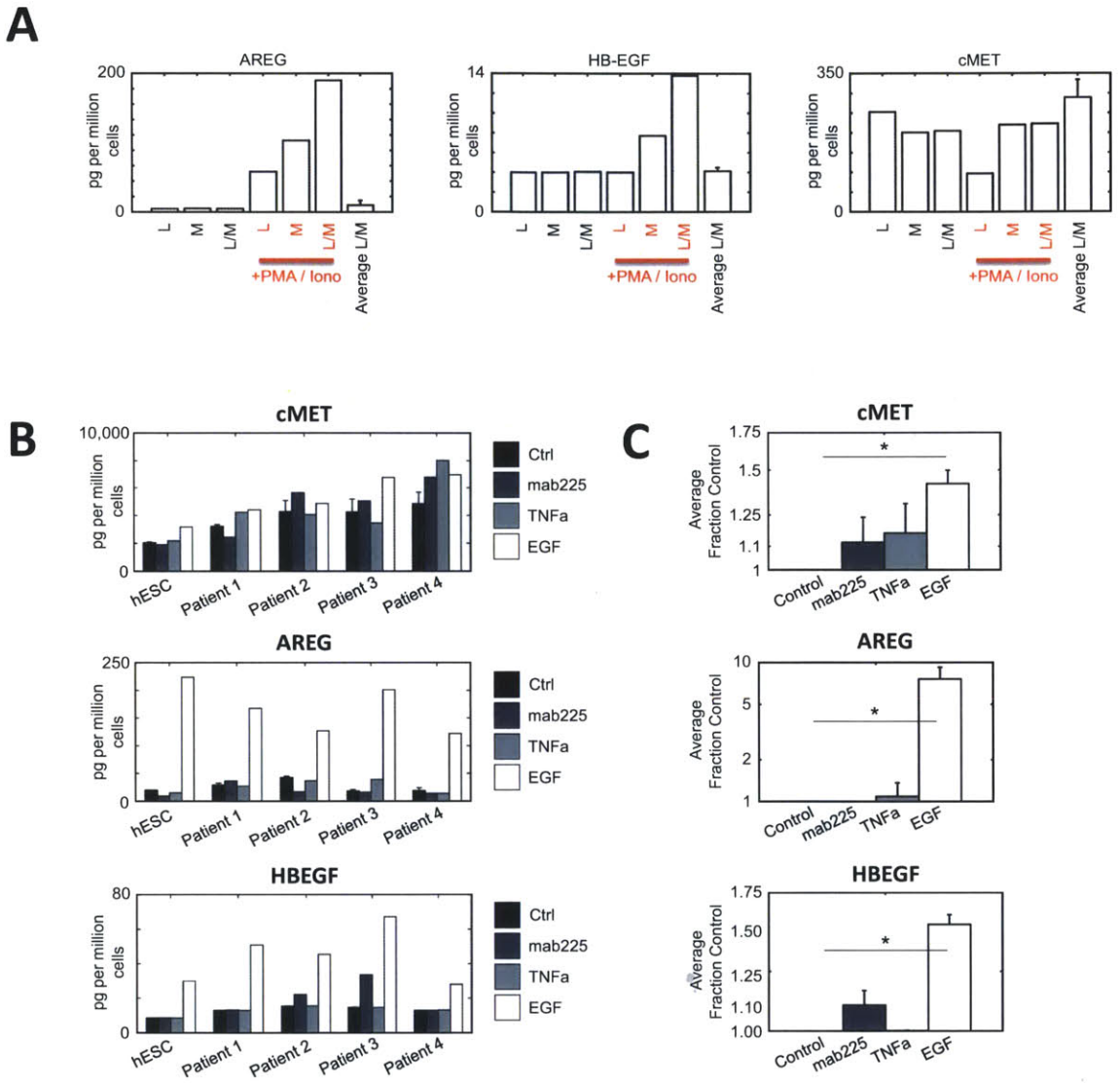

Figure C-24: **AREG, HBEGF, and MET shedding in ESC.** Telomerase-immortalized endometrial stromal cells (tHESC) and primary endometrial stromal cultures from four patients were assayed for ligand and receptor shedding 24hrs after stimulation. Results were normalized to cell count and supernatant volume, and correspond to picogram per million cells per day of analyte release (left graphs). Shedding induction was calculated relative to basal levels and averaged across all five samples (right graphs). Asterisks indicate $p<0.05$, according to the paired Student's t-test.

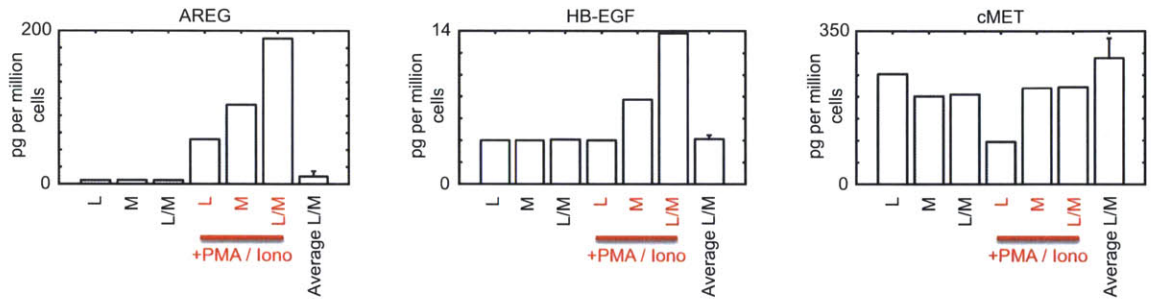

Figure C-25: **AREG, HBEGF, and MET shedding in PFMC.** Total PFMC were separated into adherent, mature monocyte populations (M) and non-adherent lymphocyte populations (L). Cells were stimulated with phorbol ester (PMA) and ionomycin (Iono). Supernatant was collected at 24hrs post-stimulation and analyzed by ELISA. "Average L/M" describes the average measurements from PFMC across nine separate patient samples, +/- standard deviation. Results were roughly normalized to cell count and supernatant volume, and correspond to picogram per million cells per day of analyte release.

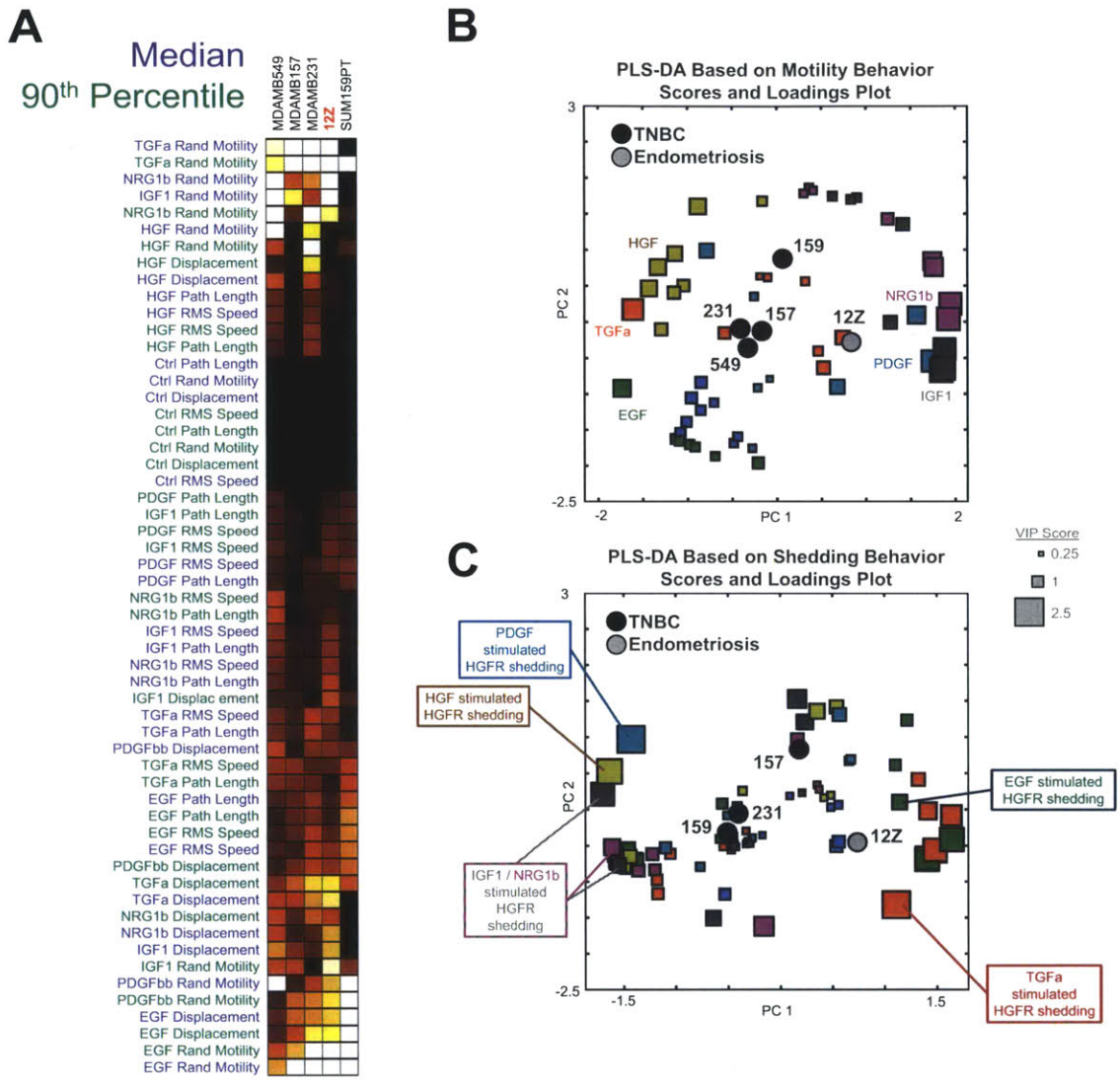

Figure C-26: **PLS-DA comparison between endometriotic and TNBC cell lines.** (A) Clustergram showing features of cell motility response to growth factors in 12Z & TNBC cells. Color indicates percent change over unstimulated conditions (black=0, white=100% increase). (B-C) PLS-DA distinguishes 12Z from TNBC cells based on motility (B) and ectodomain shedding (C) responses. Squares represent a PLS-DA loading of an individual motility (B) or shedding (C) response. Colors indicate growth factor treatment, and circles correspond to cell-line scores. Compared to TNBC cell lines, 12Z respond relatively more to NRG1b and IGF1, and relatively less to HGF. TGFa and EGF induce more MET shedding in the 12Z compared to TNBC cells. All data were mean-centered and variance normalized for each cell line and analyte.

C.1 Materials and Methods.

Growth factors and inhibitors: Recombinant growth factors and cytokines were purchased from Peprotech (Rocky Hill, NJ). For all experiments, EGF was used at a final concentration of 100ng/mL, NRG1β was used at 80ng/mL, and all others were used at 50ng/mL. Inhibitors used in this paper, their final concentrations, and vendor source are as follows: BB94 (Metalloproteinase inh.; 10uM; Tocris Bioscience), mAb225 (EGFR blocking mAb; 10ug/ml; purified from the ATCC hybridoma), gefitinib (EGFR inh.; 1uM; LC Labs), lapatinib (EGFR/HER2 inh.; 1uM; LC Labs), SP600125 (Jnk inh.; 20uM; LC Labs), TCS-6o (Jnk inh.; 10uM; Tocris), Jnk-IN-8 (Jnk inh.; 3uM; generously provided by the Gray Lab, Harvard Medical School), SB203580 (P38 inh.; 20uM; LC Labs), LY294002 (PI3K inh.; 10uM; LC Labs), U0126 (Mek inh.; 10uM; LC Labs), AZD6244 (Mek inh.; 5uM; Selleck Chem.), PD0325901 (Mek inh.; 10uM; Tocris), Foretinib (MET inh.; 100nM; Selleck), α-AREG mAb (AREG inh.; 10ug/ml; R&D Systems), pro-ADAM-9 (ADAM-9 inh.; 10uM; Biozyme, Inc.; Apex, NC), pro-ADAM-10 (ADAM-10 inh.; 4uM; Biozyme, Inc.).

Tissue culture: The 12Z cell line was generously provided by Anna Starzinski-Powitz (University of Frankfurt) by way of Steve Palmer (EMD Serono). Telomerase-immortalized human endometrial fibroblasts (tHESC CRL-4003; ATCC), SUM149PT (Asterand; Detriot, MI), MDAMB231 (ATCC), and MDAMB157 (ATCC) were cultured according to supplier recommendations. Cells were 12Z were routinely cultured in media that consisted of DMEM supplemented with 100 U/ml penicillin, 100μg/ml streptomycin (Invitrogen), along with 10% fetal bovine serum (Atlanta Biologicals; Atlanta, GA) at 37°C, 5% CO_2.

Immunoassays: Phospho-protein levels were measured using the following bead-based ELISA kits (Bio-Rad; Hercules, CA): p-Akt(Ser473), p-cJun(Ser63), p-ERK1/2(Thr202/Tyr204, Thr185/Tyr187), p-HSP27(Ser78), p-Jnk(Thr183/Tyr185), p-p38(Thr180/Tyr182), p-Src(Tyr416), p-STAT3(Tyr705), p-GSK3a/b(Ser21/Ser9), p-Tyk2(Tyr1054/Tyr1055),

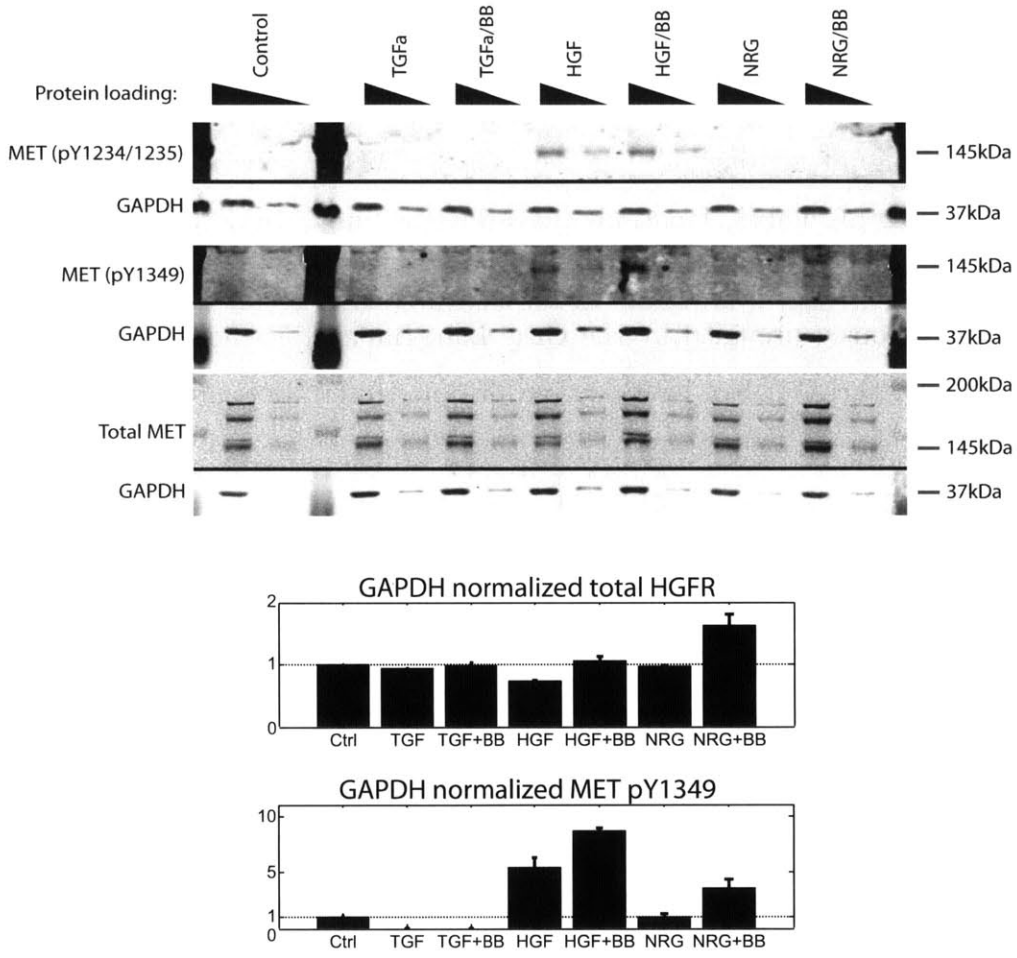

Figure C-27: **Gel images for quantification of MET shedding and phosphorylation.** Representative western blots show total and phosphorylated MET at pY1234/35 and pY1349. Densitometric quantification of results from these blots, averaged with an additional replicate blot, are shown at bottom. BB94 did not significantly change pY1234/34 levels, in contrast to pY1349 levels, which increased with BB94 treatment in the presence of NRG1b and HGF.

and p-EGFR(Tyr1173). The site-specific p-EGFR(Tyr1173) Ab was used only in Fig. 4-5C. Supernatant ligand, receptor, MMP, and TIMP measurements were performed using R&D Duo-set ELISA kits (R&D Systems, Minneapolis, MN), Widescreen bead-based ELISAs from EMD4Biosciences (Merck KGaA, Darmstadt, Germany), and Fluorokine MAP Multiplex kits (R&D Systems). EGFR, HER-2, and HER-4 phosphorylation levels were measured using a bead-conjugated ectodomain capture antibody and a pan-phosphotyrosine detection antibody (EMD4Biosciences). p-MET western blots used clone D26 for p-MET(Tyr1234/1235) and clone 130H2 for p-MET(Tyr1349), both from Cell Signaling Technologies (Danvers, MA). ADAM-17 western blots used a rabbit polyclonal antibody (Ab2051; Abcam). ADAM-10 western blots used a rabbit polyclonal antibody targeting amino acids 732-748 (Sigma). Western blots were imaged using an Odyssey (Li-cor) infrared scanner and dye-conjugated secondary antibodies (Invitrogen). An ADAM-10 bead-based ELISA was developed using an ectodomain-targeting monoclonal mouse IGG2b antibody (clone 163003, R&D Systems) for capture and a biotinylated polyclonal goat antibody (Ab936, R&D Systems) for detection. Antibodies were conjugated to carboxylated polysytrene beads (Luminex; Austin, TX) using EDC/NHS crosslinking (Thermo Fisher; Rockford, IL) following standard methods. The resulting bead-based ADAM-10 ELISA used protocols and buffers from the Bio-Rad phosphoprotein kits.

Peritoneal fluid sample and analysis: Peritoneal fluid, primary HESC, and PFMC samples were all from patients who provided informed consent in accordance with a protocol approved by the Partners Human Research Committee and the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects. We limited enrollment to pre-menopausal women with regular cycles (26-32 days), and excluded subjects having received hormone treatment within three months of surgery. Moderate/Severe (Stage III/IV) endometriosis was laparoscopically diagnosed based on the revised criteria of the American Society for Reproductive Medicine (Canis et al., 1996). Peritoneal fluid was aspirated during laparoscopy from the rectouterine pouch, following trocar insertion and before lavage or surgical ma-

nipulation. Specimens were immediately clarified within 15 min. by centrifugation, aliquoted, and stored at -80°C until further analysis.

Primary and immortalized HESC: Endometrial tissue was obtained from pipelle uterine biopsies of normally cycling pre-menopausal women. Isolation and purification of endometrial stromal fibroblasts was performed as previously described (Osteen et al., 1989). Briefly, tissue was dissected into approximately $1mm^3$ fragments using a sterile scalpel blade, transferred into a tube containing fresh complete media and centrifuged (400xg) to eliminate excess blood and debris. Fragments were resuspended in an enzyme mix containing 0.5% collagenase type IV and 0.02% deoxyribonuclease I, and 2% chicken serum in phenol red-free DMEM-F12. Suspensions were incubated at 37°C for 1hr, intermittently aspirated through decreasing sizes of glass pipettes for cell dispersion, and finally filtered through a 100um and subsequently 70um nylon cell strainer.

PFMC isolation and media conditioning: Peritoneal fluid mononuclear cells were isolated from fresh peritoneal aspirates by centrifugation (10 min. at 1000xg) and cryopreserved in complete media supplemented with 10% DMSO. Upon thawing, cells were washed and seeded at 100,000 cells/cm^2 in 24-well plates for selective adherence of mature monocytes. Non-adherent populations were removed by gentle washing after three hours of culture, and conditioned media collected following and additional 24 hours of monocyte-enriched culture. Flow cytometry routinely indicated >99% CD45 expression in fresh PFMC suspensions and >90% lymphocyte depletion in monocyte-enriched preparations.

Supernatant analysis: For quantification of supernatant analytes (Figures 1A, 2A, 3D-E, 4D, 4K, S1B, S6, S8, S9, S20), cells were plated on polystyrene plates (Corning) at 80% confluency, and stimulated the following day with serum-free media supplemented with growth factors after a 30min. pre-treatment with inhibitor. 12Z were serum starved overnight before treatment. Supernatant was collected 24hrs after stimulation, clarified by centrifugation (5min., 300g), and frozen at -20°C for

storage. At the time of supernatant collection, cells were trypsinized and analyzed for cell count and viability using ViCell instrumentation (Beckman Coulter; Brea, CA). Final analyte concentration measurements were normalized to cell count.

siRNA knockdown protocol: siRNA treatments used ON-TARGETplus SMART-pool siRNAs (Thermo Scientific), with siGENOME non-targeting siRNA pool-2 as the negative control. For both 12Z and MDAMB231, 0.5 million cells were seeded in 10 cm dishes. The following day the cells were transfected using 5 uL Dharmafect 4 and 125 pmol siRNA according to the manufacturer's protocol. One day after transfection, cells were reseeded for knockdown experiments, and 48 hrs after transfection cells were treated and lysed. Experiments for measuring ectodomain shedding and cell invasion began with growth-factor stimulation 48 hrs post-transfection, and ended 24 hrs later.

Migration assay computation: To interpret live-cell migration experiments, the root-mean-squared cell speed was calculated from position intervals between time points for each cell track, as well as the standard deviation of the mean (Kim et al., 2008). In this work, persistence denotes net displacement divided by the total path length. Endpoint invasion assays were interpreted using a modified spot finding algorithm (Santella et al., 2010) in Matlab. Briefly, confocal z-stacks were first pre-processed using background subtraction to eliminate uneven microscope illumination, followed by top-hat filtering and contrast enhancement. Pre-processed images were segmented using previously described software (Santella et al., 2010). Identified nuclei positions were assessed by nearest-neighbor analysis and PCA to infer the well-bottom where the majority of cells reside, and z-positions of the nuclei were determined as a distance from the xy-plane of the well-bottom. Unless otherwise stated, invasion metrics were calculated as a fraction of cells that had invaded further than 20um from the well-bottom, although this threshold slightly changed depending on cell-type and day-to-day variability to maximize signal-to-noise and minimize background.

Bayesian network analysis: We employed Bayesian network analysis to produce graphical representations of the conditional probabilistic dependencies among signaling events, protease activity measurements, ectodomain shedding, and cellular motility. The directed acyclic graphs produced by Bayesian network inference are comprised of two components: nodes and edges. In this application, nodes describe the various biochemical and phenotypical measurements that describe 12Z response to growth factor stimulation. The edges represent conditional probabilistic relationships between these measurements. Bayesian network inference allows for known directional relationships to be pre-specified. In this application, biochemical relationships between phosphosignaling, ADAM activity, and ectodomain shedding, have some known directionality. For example, because ADAMs are known to directly shed ectodomains, while soluble ectodomains generally do not directly influence ADAM activity, we model the relationship as directional. Furthermore, observed correlation with cell motility (Fig. 4-2A) suggests a hierarchical relationship, and is greatest on average with ectodomain shedding measurements, followed by ADAM activities, and least so for phospho-protein levels.

We use a previously described Bayesian network inference method (Eaton et al., 2007). Briefly, we derived consensus directed graph structures from exact Bayesian network model averaging over all directed acyclic graph (DAG) structures for a given number of parents per node. We averaged these results across the two algorithm parameters: the number of maximum parent edges per node (ranging between 2 and 4), and the number of k-means clusters used for data discretization (between 3 and 5). For interpretable visualization, the consensus network (Fig. 4-2C) only includes those edges with a score >0.24, which denotes that >24% of the derived Bayesian networks tested over all iterations of optimization included that edge. Although our implementation of Bayesian inference cannot capture cyclical feedback, bi-directional edges can be observed with model averaging when using a significance threshold below 0.5. In Fig. 4-2C, edge weights and color correspond to the edge score, and node color corresponds to the summed edge-weights emanating from that node.

Partial least squares analysis: Partial least squares regression (PLSR), partial least squares discriminant analysis (PLS-DA), principal component analysis (PCA), hierarchical biclustering, and all other statistical analyses were performed using Matlab (Natick, MA). Unless otherwise stated, all input and output variables were mean-centered and variance-normalized across the set of environmental stimuli, prior to PLSR, PLS-DA, or PCA. For PLSR and PLS-DA, we implemented a forward-variable selection procedure to heuristically select the minimal combination of input variables that optimally described output variable response. We iteratively added input variables to the PLSR model if they improved model fit as determined by leave-one-out cross validation accuracy. To avoid local optima, we conducted heuristic searches with several cost functions of the cross-validation accuracy, including Q^2 (the R^2 coefficient of determination for prediction accuracy), Spearman rank correlation, Pearson correlation coefficients, and for PLS-DA, area under the R.O.C. curve (AUROC). To avoid over-fitting, we required the added input variables to demonstrate PLSR loadings of greater magnitude than their observed loading standard error. Variable importance in the projection (VIP) statistics were calculated in the usual manner. Standard error for scores, loadings, and VIP were calculated by the jack-knife (Efron et al., 1983). Variance of the prediction values were computed according to Höskuldsson (Hoskuldsson et al., 1988). All results with non-zero Q^2 or AUROC accuracy demonstrated a statistical significance ($p<0.05$) as determined by the permutation test. Briefly, we randomly shuffled output variables relative to the model inputs, and ran the variable-selection procedure for each of 1000 random permutations. Shuffled data yielded a lower Q^2 value than that achieved by the actual data-set more than 95% of the time. For PLSR prediction of ADAM-10 activity in Fig. C-5, we investigated enrichment for the inclusion of phospho-proteins among PLSR models that exhibited at least 60% prediction accuracy ($Q^2 \geq 0.6$). Enrichment calculations used a hypergeometric distribution with Bonferroni correction.

Bibliography

Abate AR, Chen CH, Agresti JJ, Weitz DA (2009) Lab Chip 9: 2628-31.

Abate AR, Hung T, Mary P, Agresti JJ, Weitz DA (2010) Proc Natl Acad Sci U S A 107: 19163-19166.

Acar M, Mettetal JT, van Oudenaarden A (2008) Nat Genet 40: 471-5.

Adrain C, Zettl M, Christova Y, Taylor N, Freeman M (2012) Science 335:225-228.

Aghajanova L, Tatsumi K, Horcajadas JA, Zamah AM, Esteban FJ, et al. (2011) Biol Reprod 84: 801-815.

Agresti JJ, Antipov E, Abate AR, Ahn K, Rowat AC, et al. (2010) Proc Natl Acad Sci U S A 107: 4004-4009.

Akavia UD, Litvin O, Kim J, Sanchez-Garcia F, Kotliar D, et al. (2010) Cell 143:1005-17.

Amon LA, Law W, Fitzgibbon MP, Gross JA, O'Briant K, et al. (2010) PloS ONE 5: e11137.

Amour A, Knight CG, English WR, Webster A, Slocombe PM, et al. (2002) FEBS Lett 524: 154-158.

Appleyard DC, Chapin SC, Doyle PS (2011) Anal Chem 83:193-9.

Alvarez-Iglesias M, Wayne G, O'Dea K, Amour A, Takata M (2005) Lab Invest 85: 1440-1448.

auf dem Keller U, Bellac CL, Li Y, Lou Y, Lange PF, et al. (2010) Cancer Res 70: 7562-7569.

auf dem Keller U, Prudova A, Eckhard U, Fingleton B, Overall CM (2013) Sci Signal 6: rs2.

Banu SK, Lee JH, Starzinski-Powitz A, Arosh JA (2008) Fertil Steril 90:972-987.

Bar-Joseph Z, Gitter A, Simon I (2012) Nat Rev Gen 13:552-64.

Becker CM, Louis G, Exarhopoulos A, Mechsner S, Ebert AD, et al. (2010) Fertil Steril 94: 2343-2346.

Bedaiwy MA, Falcone T, Sharma RK, Goldberg JM, Attaran M, et al. (2002) Hum Reprod 17: 426-431.

Belsley D (1991) Wiley, New York.

Berezhkovskii AM, Coppey M, Shvartsman SY (2009) Proc Natl Acad Sci U S A 106: 1087-92.

Bhargava R, Gerald WL, Li AR, Pan Q, Lal P, et al. (2005) Mod Pathol 18: 1027-1033.

Blobel CP. (2005) Nat Rev Mol Cell Biol 6: 32-43.

Blum G, Mullins SR, Keren K, Fonovic M, Jedeszko C, et al. (2005) Nat Chem Biol 1: 203-209.

Bodenmiller B, Zunder ER, Finck R, Chen TJ, Savig ES, et al. (2012) Nat Biotech 30:858-67.

Bouchier-Hayes L, Muñoz-Pinedo C, Connell S, Green D (2008) Methods 44: 222-228.

Brouzes E, Medkova M, Savenelli N, Marran D, Twardowski M, et al. (2009) Proc Natl Acad Sci U S A 106: 14195-14200.

Brown CL, Meise KS, Plowman GD, Coffey RJ, Dempsey PJ (1998) J Biol Chem 273: 17258-17268.

Bruner-Tran KL, Eisenberg E, Yeaman GR, Anderson TA, McBean J, et al. (2002) J Clin Endocrinol Metab 87: 4782-4791.

Bruner KL, Matrisian LM, Rodgers WH, Gorstein F, Osteen KG. (1997) J Clin Invest 99: 2851.

Bugge TH, Antalis TM, Wu Q (2009) J Biol Chem 284: 23177-81.

Butler GS, Dean RA, Morrison CJ, Overall CM (2010) Methods Mol Biol 622: 451-470.

Caescu C, Jeschke G, Turk B (2009) Biochem J 424: 79-88.

Cancer Genome Atlas Network (2012) Nature 490: 61-70.

Canis M, Donnez JG, Guzick DS, Halme JK, Rock JA, et al. (1997) Fertil Steril 67: 817-821.

Carey LA, Rugo HS, Marcom PK, Mayer EL, Esteva FJ, et al. (2012) J Clin Oncol 30: 2615-2623.

Carregal-Romero S, Caballero-Diaz E, Beqa L, Abdelmonem AM, et al. (2013) Annu Rev Anal Chem (Palo Alto Calif). Feb 28 [Epub ahead of print]

Cayan F, Ertunc D, Aras-Ates N, Ayaz L, Akbay E, et al. (2010) Fertil Steril 94: 1622-1626.

Chan KM, Wong HL, Jin G, Liu B, Cao R, et al. (2012) Cell 22: 1176-90.

Chen B, Platt MO (2011) J Transl Med 9: 109. (*" Chen and Platt, 2011"*)

Chen CH, Miller MA, Sarkar A, Beste MT, Isaacson KB, et al. (2012) J Am Chem Soc 135: 1645-1648.

Chen CH, Sarkar A, Song YA, Miller MA, Kim SJ, et al. (2011) J Am Chem Soc 133: 10368-10371. (*" Chen et al., 2011"*)

Chen E, Kridel S, Howard E, Li W, Godzik A, Smith JW (2002) J Biol Chem 277: 4485-4492.

Chen WW, Schoeberl B, Jasper PJ, Niepel M, Nielsen UB, et al. (2009) Mol Syst Biol 5:1-19.

Chung HW, Lee JY, Moon HS, Hur SE, Park MH, et al. (2002) Fertil Steril 78: 787-795.

Chung HW, Wen Y, Chun SH, Nezhat C, Woo BH, et al. (2001) Fertil Steril 75: 152-159.

Cinar O, Seval Y, Uz YH, Cakmak H, Ulukus M, et al. (2009) Reprod Biomed Online 19: 864-871.

Clark JC, Thomas DM, Choong PF, Dass CR (2007) Cancer Metastasis Rev 26:675-83.

Collette T, Maheux R, Mailloux J, Akoum A (2006) Hum Reprod 21: 3059-3067.

Corin KA, Gibson LJ (2010) Biomaterials 31: 4835-45.

Cornish-Bowden A (1995) Portland, London, 3rd edn.

Coticchia CM, Curatolo AS, Zurakowski D, Yang J, Daniels KE, et al. (2011) Gynecol Oncol 123:295-300.

Cukierman E, Pankov R, Stevens DR, Yamada KM (2001) Science 294: 1708-1712.

Curry TE Jr, Osteen KG (2003) Endocr Rev 24: 428-65.

Dang M, Dubbin K, D'Aiello A, Hartmann M, Lodish H, Herrlich A (2011) J Biol Chem 286: 17704-13.

De Leon FD, Vijayakumar R, Brown M, Rao CV, Yussman MA, et al. (1986) Obstet Gynecol 68: 189-194.

Dean R, Overall C. (2007) Mol Cell Proteomics 6: 611-623.

Decock J, Thirkettle S, Wagstaff L, Edwards DR (2011) J Cell Mol Med 15: 1254-65.

DeForest CA, Anseth KS (2012) Annu Rev Chem Biomol Eng 3: 421-44.

Deryugina E, Quigley J (2006) Canc Met Rev 25: 9-34.

Deu E, Verdoes M, Bogyo M (2012) Nat Struct Mol Biol 19: 9-16.

Devy L, Dransfield DT (2011) Biochem Res Int 2011: 191670.

Doedens JR, Mahimkar RM, Black RA (2003) Biochem Biophys Res Commun 308: 331-338.

Dorman G, Cseh S, Hajdu I, Barna L, Konya D, et al. (2010) Drugs 70: 949-964.

Drag M, Bogyo M, Ellman J, Salvesen G (2010) J Biol Chem 285: 3310-3318.

Duffy M, McKiernan E, O'Donovan N, McGowan P (2009) Clin Canc Res 15: 1140-1144.

Dufour A, Overall CM (2013) Trends Pharmacol Sci Mar 26 doi: 10.1016/j.tips.2013.02.004.

Duncan JS, Whittle MC, Nakamura K, Abell AN, Midland AA, et al. (2012) Cell 149: 307-321.

Efron B, Gong G (1983) Am Stat 37: 36-48.

Egeblad M, Werb Z (2002) Nat Rev Canc 2: 163-176.

Eisenberg MC, Kim Y, Li R, Ackerman WE, Kniss DA, Friedman A (2011) Proc Natl Acad Sci U S A 108: 20078-83.

Ejskjaer K, Sorensen BS, Poulsen SS, Mogensen O, Forman A, Nexo E (2009) Gynecol Obstet Invest 67:118-126.

El-Ghobashy AA, Shaaban AM, Innes J, Prime W, Herrington CS (2005) Int J Gynecol Cancer 15: 503-509.

Eming S, Krieg T, Davidson J (2007) J Invest Dermatol 127: 514-525.

Entenberg D, Kedrin D, Wyckoff J, Sahai E, Condeelis J, Segall JE (2013) Curr Protoc Cell Biol 19: 19.7.

Felber LM, Borgoo CA, Cloutier SM, Kndig C, Kishi T, et al. (2005) J Biol Chem 386: 291-298.

Finas D, Huszar M, Agic A, Dogan S, Kiefel H, et al. (2008) Hum Reprod 23: 1053-1062.

Fingleton B. (2008) Sem Cell Dev Biol 19: 61-68.

Foveau B, Ancot F, Leroy C, Petrelli A, Reiss K, et al. (2009) Mol Biol Cell 20: 2495-2507.

Fowlkes J, Thrailkill K, Serra D, Nagase H (1997) Endocrine 7: 33-36.

Fraley SI, Feng Y, Krishnamurthy R, Kim DH, Celedon A, et al. (2010) Nat Cell Biol 12: 598-604.

Fridman JS, Caulder E, Hansbury M, Liu X, Yang G, et al. (2007) Clin Canc Res 13: 1892-902.

Friedl P, Sahai E, Weiss S, Yamada KM (2012) Nat Rev Mol Cell Biol 13: 743-7.

Gaggioli C, Hooper S, Hidalgo-Carcedo C, Grosse R, Marshall JF, et al. (2007) Nat Cell Biol 9: 1392-400.

Galis Z, Sukhova G, Libby P (1995) FASEB J 9: 974-980.

Ghedini GC, Ciravolo V, Tortoreto M, Giuffr S, Bianchi F, et al. (2010) J Cell Physiol 225: 256-265.

Gilabert-Estelles J, Estelles A, Gilabert J, Castello R, Espana F, et al. (2003) Hum Reprod 18: 1516-1522.

Gill SE, Kassim SY, Birkland TP, Parks WC (2010) Methods Mol Biol 622: 31-52.

Gobaa S, Hoehnel S, Roccio M, Negro A, Kobel S, Lutolf MP (2011) Nat Methods 8: 949-55.

Gooz P, Dang Y, Higashiyama S, Twal WO, Haycraft CJ, Gooz M (2012) PLoS ONE 7: e33350.

Gough W, Hulkower KI, Lynch R, McGlynn P, Uhlik M, et al. (2011) J Biomol Screen 16: 155-63.

Gottschal C, Malberg K, Arndt M, Schmitt J, Roessner A, et al. (2000) Adv Exp Med Biol 477:483-486.

Gosalia D, Diamond SL (2010) Methods Mol Biol 669: 69-78.

Gosalia D, Salisbury C, Maly D, Ellman J, Diamond S (2005) Proteomics 5: 1292-1298.

Gosalia D, Denney W, Salisbury C, Ellman J, Diamond S (2006) Biotechnol Bioeng 94: 1099-1110.

Goswami S, Sahai E, Wyckoff JB, Cammer M, Cox D, et al. (2005) Cancer Res 65: 5278-5283.

Grund EM, Kagan D, Tran CA, Zeitvogel A, Starzinski-Powitz A, et al. (2008) Mol Pharmacol 73: 1394-1404.

Gueders M, Foidart J, Noel A, Cataldo D (2006) Eur J Pharmacol 533: 133-144.

Guiet R, Van Goethem E, Cougoule C, Balor S, Valette A, et al. (2011) J Immunol 187: 3806-3814.

Guo MT, Rotem A, Heyman JA, Weitz DA (2012) Lab Chip 12: 2146-55.

Gurgan T, Bukulmez O, Yarali H, Tanir M, Akyildiz S (1999) J Reprod Med 44: 450-454.

Haass-Koffler CL, Naeemuddin M, Bartlett SE (2012) J Vis Exp Aug 31: e4233.

Hadler-Olsen E, Kanapathippillai P, Berg E, Svineng G, Winberg JO, Uhlin-Hansen L (2010) J Histochem Cytochem 58: 29-39.

Hall KC, Blobel CP (2012) PLoS ONE 7: e31600.

Han J, Chang H, Giricz O, Lee GY, Baehner FL, et al. (2010) PLoS Comput Biol 26: e1000684.

Handsley M, Edwards D (2005) Int J Cancer 115: 849-860.

Harjanto D, Zaman MH (2010) Ann Biomed Eng 38: 1815-25.

Haro H, Crawford H, Fingleton B, Shinomiya K, Spengler D, Matrisian LM (2000) J Clin Invest 105: 143-150.

Hartmann D, de Strooper B, Serneels L, Craessaerts K, Herreman A, et al. (2002) Hum Mol Genet 11: 2615-2624.

Hassemer EL, Le Gall SM, Liegel R, McNally M, Chang B, et al. (2010) Genetics 185: 245-255.

He F, Chen H, Probst-Kepper M, Geffers R, Eifes S, et al. (2012) Mol Syst Biol 8: 624.

Hernandez-Barrantes S, Toth M, Bernardo MM, Yurkova M, Gervasi DC, et al. (2000) J Biol Chem 275: 12080-12089.

Herrlich A, Klinman E, Fu J, Sadegh C, Lodish H (2008) FASEB J 22: 4281-4295.

Hidalgo-Carcedo C, Hooper S, Chaudhry SI, Williamson P, Harrington K, et al. (2011) Nat Cell Biol 13: 49-58.

Hill SM, Lu Y, Molina J, Heiser LM, Spellman PT, et al. (2012) Bioinformatics 28: 2804-10.

Hiraoka Y, Yoshida K, Ohno M, Matsuoka T, Kita T, Nishi E (2008) Biochem Biophys Res Commun 370: 154-158.

Hoeflich KP, O'Brien C, Boyd Z, Cavet G, Guerrero S, et al. (2009) Clin Canc Res 15: 4649-4664.

Honda H, Barrueto FF, Gogusev J, Im DD, Morin PJ (2008) Reprod Biol Endocrinol 6: 59.

Horiuchi K, Le Gall S, Schulte M, Yamaguchi T, Reiss K, et al. (2007) Substrate selectivity of epidermal growth factor-receptor ligand sheddases and their regulation by phorbol esters and calcium influx. Mol Biol Cell 18: 176-188.

Höskuldsson A (1988) J Chemometr 2: 211-228.

Huang C, Rajfur Z, Borchers C, Schaller MD, Jacobson K. (2003) Nature 424(6945):219-223.

Huang HF, Hong LH, Tan Y, Sheng JZ (2004) Fertil Steril 81: 1235-1239.

Huang JC, Papasakelariou C, Dawood MY (1996) Fertil Steril 65: 931-934.

Huang SC, Fraenkel E (2009) Sci Signal 2:ra40.

Hughes-Alford SK, Lauffenburger DA (2012) Curr Opin Cell Biol 24: 284-91.

Hyduke DR, Palsson BO (2010) Nat Rev Genet 11: 297-307.

Ideker T, Lauffenburger D (2003) Trends Biotechnol 21: 255-262.

Impens F, Colaert N, Helsens K, Plasman K, van Damme P, et al. (2010) Proteomics 10: 1284-96.

Inagaki M, Yoshida S, Kennedy S, Takemura N, Deguchi M, et al. (2007) Gynecol Endocrinol 23(8):474-478.

Inoue A, Ishiguro J, Kitamura H, Arima N, Okutani M, et al. (2012) Nat Methods 9: 1021-9.

Irvine DJ, Mayes AM, Griffith LG (2001) Biomacromolecules 2: 85-94.

Itoh F, Komohara Y, Takaishi K, Honda R, Tashiro H, et al. (2013) Mol Cell Endocrinol doi: 10.1016/j.fertnstert.2013.01.133. [Epub ahead of print].

Janes KA, Yaffe MB (2006) Nat Rev Mol Cell Biol 7: 820-828.

Jefferson T, Auf dem Keller U, Bellac C, Metz VV, Broder C, et al. (2013) Cell Mol Life Sci 70: 309-33.

Joslin E, Opresko L, Wells A, Wiley HS, Lauffenburger D (2007) J Cell Sci 120: 3688-3699.

Joslin E, Shankaran H, Opresko L, Bollinger N, Lauffenburger D, Wiley HS (2010) Mol BioSyst 6: 1293-1306.

Kalu E, Sumar N, Giannopoulos T, Patel P, Croucher C, et al. (2007) J Obstet Gynaecol Res 33: 490-495.

Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, et al. (2012) Cell 150: 389-401.

Katsu M, Niizuma K, Yoshioka H, Okami N, Sakata H, Chan PH (2010) J Cereb Blood Flow Metab 30: 1939-50.

Kemp ML, Wille L, Lewis CL, Nicholson LB, Lauffenburger DA (2007) J Immunol 178: 4984-92.

Keow JY, Pond ED, Cisar JS, Cravatt BF, Crawford BD (2012) PLoS One 7: e43434.

Kessenbrock K, Plaks V, Werb Z (2010) Cell 141: 52-67.

Khan KN, Masuzaki H, Fujishita A, Kitajima M, Sekine I, et al. (2003) Fertil Steril 79: 173-181.

Kim HD, Guo TW, Wu AP, Wells A, Gertler FB, Lauffenburger DA (2008) Mol Biol Cell 19:4249-4259.

Kim HD, Meyer AS, Wagner JP, Alford SK, Wells A, et al. (2011) Mol Cell Proteomics 10: M111.008433.

Kirfel G, Rigort A, Borm B, Herzog V (2004) Eur J Cell Biol 83: 717-724.

Kleifeld O, Doucet A, Prudova A, auf dem Keller U, Gioia M, et al. (2011) Nat Protoc 6: 1578-1611.

Kleiman LB, Maiwald T, Conzelmann H, Lauffenburger DA, Sorger PK (2010) Mol Cell 43: 723-37.

Kleiner D, Stetler-Stevenson W (1994) Anal Biochem 218: 325-329.

Koga K, Osuga Y, Tsutsumi O, Okagaki R, Momoeda M, et al. (2000) Mol Hum Reprod 6: 929-933.

Kukar TL, Ladd TB, Bann MA, Fraering PC, Narlawar R, et al. (2008) Nature 453: 925-9.

Kveiborg M, Instrell R, Rowlands C, Howell M, Parker PJ (2011) PLoS ONE 6: e17168.

Kyama CM, Overbergh L, Debrock S, Valckx D, Vander Perre S, et al. (2006) Fertil Steril 85: 1667-1675.

Lau KS, Juchheim AM, Cavaliere KR, Philips SR, Lauffenburger DA, Haigis KM (2011) Sci Signal 4: ra16.

Laudanski P, Szamatowicz J, Kowalczuk O, Kuzmicki M, Grabowicz M, et al. (2009) Hum Reprod 24: 1880-1890.

Lauer-Fields JL, Nagase H, Fields GB (2004) J Biomol Tech 15: 305-16.

Lauffenburger DA, Linderman JJ (1996) Receptors: models for binding, trafficking, and signaling. Oxford University Press on Demand.

Lawson CL, Hanson RJ (1995). Solving least squares problems: 15.

Le Gall SM, Bob P, Reiss K, Horiuchi K, Niu XD, et al. (2009) Mol Biol Cell 20: 1785-1794.

Le Gall SM, Maretzky T, Issuree PD, Niu XD, Reiss K, et al. (2010) J Cell Sci 123: 3913-22.

Lee J, Banu SK, Subbarao T, Starzinski-Powitz A, Arosh JA (2011) Mol Cell Endocrinol 332: 306-313.

Lee MJ, Ye AS, Gardino AK, Heijink AM, Sorger PK, et al. (2012) Cell 149: 780-794.

Lembong J, Yakoby N, Shvartsman SY (2009) Proc Natl Acad Sci U S A 106: 3213-8.

Leong DJ, Hardin JA, Cobelli NJ, Sun HB (2011) Ann N Y Acad Sci 1240: 32-37.

Lim M, Craik C (2009) Bioorg Med Chem 17: 1094-1100.

Liu ET & Lauffenburger DA (2009) Systems biomedicine: concepts and perspectives. Academic Press.

Liu X, Fridman JS, Wang Q, Caulder E, Yang G, et al. (2006) Cancer Biol Ther 5: 648-656.

López-Otín C, Hunter T (2010) Nat Rev Canc 10: 278-92.

López-Otín C, Matrisian LM (2007) Nat Rev Canc 7: 800-808.

Maheshwari G, Wiley HS, Lauffenburger DA (2001) J Cell Biol 155: 1123-1128.

Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, et al. (2006) BMC Bioinformatics. 7 Suppl 1:S7.

Maretzky T, Evers A, Zhou W, Swendeman SL, Wong PM, et al. (2011) Nat Commun 2:229.

Martin NK, Gaffney EA, Gatenby RA, Maini PK (2010) J Theor Biol 267: 461-70.

Mason SD, Joyce JA (2011) Trends Cell Biol 21: 228-37.

Matsuzaki S, Canis M, Vaurs-Barrire C, Pouly JL, Boespflug-Tanguy O, et al. (2004) Mol Hum Reprod 10: 719-728.

Matalliotakis IM, Goumenou AG, Koumantakis GE, Neonaki MA, Koumantakis EE, et al. (2003) Int Immunopharmacol 3: 81-89.

McDonagh CF, Huhalov A, Harms BD, Adams S, Paragas V, et al. (2012) Mol Cancer Ther 11: 582-93.

McIlwain DR, Lang PA, Maretzky T, Hamada K, Ohishi K, et al. (2012) Science 335: 229-232.

Meacham CE, Ho EE, Dubrovsky E, Gertler FB, Hemann MT (2009) Nat Genet 41: 1133-1137.

Melin A, Sparen P, Bergqvist A (2008) Obstet Gynecol Surv 63:156.

Meyer AS, Hughes-Alford SK, Kay JE, Castillo A, Wells A, et al. (2012) J Cell Biol 197:721-729.

Miller MA, Barkal L, Jeng K, Herrlich A, Moss M, et al. (2011) Integr Biol (Camb) 3:422-438.

Miller MA, Feng XJ, Li G, Rabitz HA (2012) PLoS ONE 7: e37664. (*"Miller et al., 2012a"*)

Miller M, Hafner M, Sontag E, Davidsohn N, Subramanian S, et al. (2012) PLoS Comput Biol 8: e1002579. (*"Miller et al., 2012b"*)

Miller OJ, El Harrak A, Mangeat T, Baret JC, Frenz L, et al. (2012) Proc Natl Acad Sci U S A 109: 378-383. (*"Miller et al., 2012c"*)

Miller-Jensen K, Janes KA, Brugge JS, Lauffenburger DA (2007) Nature 448: 604-608.

Mohammed FF, Smookler DS, Taylor SE, Fingleton B, Kassiri Z, et al. (2004) Nat Genet 36: 969-977.

Mohan MJ, Seaton T, Mitchell J, Howe A, Blackburn K, et al. (2002) Biochemistry 41: 9462-9469.

Mori H, Lo AT, Inman JL, Alcaraz J, Ghajar CM, et al. (2013) Development 140: 343-52.

Morris MK, Melas I, Saez-Rodriguez J (2013) Methods Mol Biol 930: 179-214.

Morris MK, Saez-Rodriguez J, Clarke DC, Sorger PK, Lauffenburger DA (2011) PLoS Computat Biol 7: e1001099.

Morris MK, Saez-Rodriguez J, Sorger PK, Lauffenburger DA (2010) Biochemistry 49: 3216-3224.

Morrison CJ, Butler GS, Rodrguez D, Overall CM. (2009) Curr Opin in Cell Biol 21: 645-653.

Moss M, Rasmussen F (2007) Anal Biochem 366: 144-148. (*"Moss and Ramussen, 2007"*)

Moss M, Rasmussen F, Nudelman R, Dempsey P, Williams J (2009) Comb Chem High Throughput Screen 13: 358-365.

Moss ML, Bomar M, Liu Q, Sage H, Dempsey P, et al. (2007) J Biol Chem 282: 35712-35721. (*"Moss et al., 2007"*)

Moss ML, Powell G, Miller MA, Edwards L, Qi B, et al. (2011) J Biol Chem 29: 3020-3032.

Muranen T, Selfors LM, Worster DT, Iwanicki MP, Song L, et al. (2012) Cancer Cell 21: 227-239.

Murk W, Atabekoglu CS, Cakmak H, Heper A, Ensari A, et al. (2008) J Clin Endocrinol Metab 93: 3532-3540.

Murphy G (2008) Nat Rev Canc 8: 932-941.

Murthy A, Cruz-Munoz W, Khokha R (2008) Springer Verlag: 373-400.

Nagase H, Fields G (1998) Biopolymers 40: 399-416.

Naresh A, Long W, Vidal GA, Wimley WC, Marrero L, et al. (2006) Cancer Res 66: 6412-20.

Nath A, Atkins WM (2006) Drug Metab Dispos 34: 1433-1435.

Ngo C, Nicco C, Leconte M, Chereau C, Arkwright S, et al. (2010) J Pathol 222: 148-157.

Page-McCaw A, Ewald A, Werb Z (2007) Nat Rev Mol Cell Biol 8: 221-233.

Palmier MO, Van Doren SR (2007) Anal Biochem 371: 43-51.

Park KY, Li WA, Platt MO (2012) Integr Biol (Camb) 4: 1459-69.

Pathak A, Kumar S (2012) Proc Natl Acad Sci U S A 109: 10334-9.

Peluso C, Christofolini DM, Goldman CS, Mafra FA, Cavalcanti V, et al. (2012) Hum Immunol 74: 93-97.

Philippoussis F, Gagné D, Hugo P, Gosselin D (2004) J Soc Gynecol Investig 11: 175-181.

Pillinger MH, Marjanovic N, Kim SY, Scher JU, Izmirly P, et al. (2005) J Biol Chem 280: 9973-9979.

Prefumo F, Venturini PL, Fulcheri E. (2003) Int J Gynecol Pathol 22: 83-88.

Preus S, Wilhelmsson LM (2012) Chembiochem 13: 1990-2001.

Protopapas A, Markaki S, Mitsis T, Milingos D, Athanasiou S, et al. (2010) Fertil Steril 94: 2470-2472.

Obach RS, Reed-Hagen AE (2002) Drug Metab Dispos 30: 831-837.

Oberhardt MA, Palsson BO, Papin JA (2009) Mol Syst Biol 5: 320.

Ohkubo S, Miyadera K, Sugimoto Y, Ki M, Wierzba K, Yamada Y (2001) Comb Chem High Throughput Screen 4: 573-583.

Ohlsson Teague EM, Van der Hoek KH, Van der Hoek MB, Perry N, Wagaarachchi P, et al. (2009) Mol Endocrinol 23: 265-275.

Oliver G, Leferson J, Stetler-Stevenson W, Kleiner D (1997) Anal Biochem 244: 161-166.

Olson GE, Winfrey VP, Matrisian PE, NagDas SK, Hoffman LH (1998) Cell Tissue Res 293: 489-498.

Osteen KG, Hill GA, Hargrove JT, Gorstein F (1989) Fertil Steril 52: 965-972.

Osuga Y, Tsutsumi O, Okagaki R, Takai Y, Fujimoto A, et al. (1999) Hum Reprod 14: 1611-1613.

Ota I, Li XY, Hu Y, Weiss SJ (2009) Proc Natl Acad Sci U S A 106: 20318-23.

Overall CM, Blobel CP (2007) Nat Rev Mol Cell Biol 8: 245-57.

Overall CM, Dean RA (2006) Cancer Metastasis Rev 25: 69-75.

Overton C, Fernandez-Shaw S, Hicks B, Barlow D, Starkey P (1996) Hum Reprod 11: 380-386.

Ozkan S, Murk W, Arici A. (2008) Annals NY Acad of Sci 1127: 92-100.

Packard BZ, Artym VV, Komoriya A, Yamada KM (2009) Matrix Biol 28: 3-10.

Painter KJ, Armstrong NJ, Sherratt JA (2010) J Theor Biol 264: 1057-67.

Pan Q, van der Laan LJ, Janssen HL, Peppelenbosch MP (2012) Trends Biotechnol 30: 206-15.

Park KY, Li WA, Platt MO (2012) Integr Biol (Camb) 4: 1459-69.

Pe'er D, Hacohen N (2011) Cell 144: 864-73.

Petrie RJ, Gavara N, Chadwick RS, Yamada KM (2012) J Cell Biol 197: 439-455.

Pilka R, Domanski H, Hansson S, Eriksson P, Casslen B (2004) Mol Hum Reprod 10: 641-650.

Pilka R, Oborna I, Lichnovsky V, Havelka P, Fingerova H, et al. (2006) Hum Reprod 21: 3146-3156.

Pizzo A, Salmeri FM, Ardita FV, Sofo V, Tripepi M, et al. (2002) Gynecol Obstet Invest 54: 82-87.

Platt MO, Wilder CL, Wells A, Griffith LG, Lauffenburger DA (2009) Stem Cells 27: 2804-2814.

Prill RJ, Saez-Rodriguez J, Alexopoulos LG, Sorger PK, Stolovitzky G (2011) Sci Signal 4: mr7.

Primakoff P, Myles DG (2000) Trends Genet 16: 83-7.

Pregibon DC, Toner M, Doyle PS (2007) Science 315: 1393-96.

Pujol A, Mosca R, Farres J, Aloy P (2010) Trends Pharmacol Sci 31: 115-23.

Ramon L, Gilabert-Estelles J, Castello R, Gilabert J, Espana F, et al. (2005) Hum Reprod 20: 272-278.

Rantala JK, Pouwels J, Pellinen T, Veltel S, Laasola P, et al. (2011) Nat Cell Biol 13: 1315-1324.

Rasmussen F, Yeung N, Kiefer L, Murphy G, Lopez-Otin C, et al. (2004) Biochemistry 43: 2987-2995.

Rawlings, ND (2010) Biochimie 92: 1463-83.

Rawlings ND, Barrett AJ, Bateman A (2009) Nucleic Acids Res 38: D227-D233.

Ridky TW, Chow JM, Wong DJ, Khavari PA (2010) Nat Med 16: 1450-5.

Ripley D, Tunuguntla R, Susi L, Chegini N (2006) Int J Gyn Cancer 16: 1794-1800.

Robinson JA, Characklis WG (1984) Microbial Ecology 10: 165-178.

Roccio M, Gobaa S, Lutolf MP (2012) Integr Biol (Camb) 4: 391-400.

Rocks N, Paulissen G, El Hour M, Quesada F, Crahay C, et al. (2008) Biochimie 90: 369-379.

Rodriguez D, Morrison CJ, Overall CM (2010) Biochim Biophys Acta 1803: 39-54.

Roy R, Zurakowski D, Pories S, Moss ML, Moses MA (2011) Clin Biochem 44: 1434-1439.

Rundhaug J (2007) J Cell Mol Med 9: 267-285.

Sabeh F, Shimizu-Hirota R, Weiss SJ (2009)J Cell Biol 185: 11-19.

Sacan A, Ferhatosmanoglu H, Coskun H (2008) Bioinformatics 24: 1647-9.

Saez-Rodriguez J, Alexopoulos LG, Zhang M, Morris MK, Lauffenburger DA, Sorger PK (2011) Cancer Res 71: 5400-11.

Saito T, Mizumoto H, Kuroki K, Fujii M, Mori S, et al. (1995) Nippon Sanka Fujinka Gakkai Zasshi 47: 495-496.

Saghatelian A, Jessani N, Joseph A, Humphrey M, Cravatt B (2004) Proc Natl Acad Sci U S A 101: 10000-10005.

Sahin U, Weskamp G, Kelly K, Zhou H, Higashiyama S, et al. (2004) J Cell Biol 164: 769-779.

Santella A, Du Z, Nowotschin S, Hadjantonakis AK, Bao Z (2010) BMC Bioinformatics 11: 580.

Schelter F, Grandl M, Seubert B, Schaten S, Hauser S, et al. (2011) Clin Exp Metastasis 28: 793-802.

Schelter F, Kobuch J, Moss ML, Becherer JD, Comoglio PM, et al. (2010) J Biol Chem 285: 26335-40.

Schilling O, Huesgen PF, Barr O, Auf dem Keller U, Overall CM (2011) Nat Protoc 6: 111-20.

Schilling O, Overall CM (2007) Curr Opin Chem Biol 11: 36-45.

Schneider EL, Craik CS (2009) Methods Mol Biol 539: 59-78.

Shan K, Lian-Fu Z, Hui D, Wei G, Na W, et al. (2006) Mol Hum Reprod 12: 35-39.

Shankaran H, Ippolito DL, Chrisler WB, Resat H, Bollinger N, et al. (2009) Mol Syst Biol 5: 332.

Sharpe-Timms KL, Keisler LW, McIntush EW, Keisler DH (1998) Fertil Steril 69: 1128-1134.

Shazand K, Baban S, Priv C, Malette B, Croteau P, et al. (2004) Mol Hum Reprod 10: 871-877.

Sheng Q, Liu X, Fleming E, Yuan K, Piao H, et al. (2010) Cancer Cell 17: 298-310.

Sieber S, Niessen S, Hoover H, Cravatt B (2006) Nat Chem Biol 2: 274-281.

Sieuwerts AM, Meijer-van Gelder ME, Timmermans M, Trapman A, Garcia RR, et al (2005) Clin Canc Res 11: 7311-7321.

Simms JS, Chegini N, Williams RS, Rossi AM, Dunn WA Jr (1991) Obstet Gynecol 78: 850-857.

Simpson KJ, Selfors LM, Bui J, Reynolds A, Leake D, et al. (2008) Nat Cell Biol 10: 1027-1038.

Sluzky V, Klibanov AM, Langer R (1992) Biotechnol Bioeng 40: 895-903.

Stains CI, Tedford NC, Walkup TC, Lukovic E, Goguen BN, et al. (2012) Chem Biol 19: 210-217.

Stamenkovic I (2003) J Pathol 200: 448-464.

Starr AE, Dufour A, Maier J, Overall CM (2012) J Biol Chem 287: 5848-60.

Steff AM, Gagné D, Page M, Rioux A, Hugo P, et al. (2004) Am J Reprod Immunol 51: 166-173.

Sternlicht M, Werb Z (2001) Annu Rev Cell Dev Biol 17: 463-416.

Storey JD (2002) J Roy Stat Soc B 64: 479-498.

Stoter M, Niederlein A, Barsacchi R, Meyenhofer F, Brandl H, Bickle M (2013) Methods Mol Biol 986: 105-22.

Sugiyama N, Varjosalo M, Meller P, Lohi J, Chan KM, et al. (2010) Proc Natl Acad Sci U S A 107: 15786-91.

Szamatowicz J, Laudański P, Tomaszewska I (2002) Hum Reprod 17: 284-288.

Takahashi T, Naito S, Onoda J, Yamauchi A, Nakamura E, et al. (2012) Clin Chim Acta 413: 1591-9.

Tam E, Morrison C, Wu Y, Stack M, Overall C (2004) Proc Natl Acad Sci U S A 101: 6917-6922.

Tape CJ, Willems SH, Dombernowsky SL, Stanley PL, Fogarasi M, et al. (2011) Proc Natl Acad U S A 108: 5578-83.

Turke AB, Zejnullahu K, Wu YL, Song Y, Dias-Santagata D, et al. (2010) Cancer Cell 17: 77-88.

Uz YH, Murk W, Bozkurt I, Kizilay G, Arici A, et al. (2011) Histochem Cell Biol 135: 83-91.

Uzan C, Darai E, Valent A, Graesslin O, Cortez A, et al. (2009) Virchows Arch 454: 525-529.

Verspurten J, Gevaert K, Declercq W, Vandenabeele P (2009) Trends Biochem Sci 34: 319-323.

Villanueva J, Vultur A, Lee JT, Somasundaram R, Fukunaga-Kalabis M, et al. (2010) Cancer Cell 18: 683-95.

Vu T, Werb Z (2000) Genes Dev 14: 2123-2133.

Wagner EF, Nebreda AR (2009) Nat Rev Canc 9: 537-549.

Walker E, Rosenberg G (2009) Exp Neurol 216: 122-131.

Wilson TR, Fridlyand J, Yan Y, Penuel E, Burton L, et al. (2012) Nature 487: 505-9.

Witters L, Scherle P, Friedman S, Fridman J, Caulder E, et al. (2008) Cancer Res 68: 7083-7089.

Wolf K, Mazo I, Leung H, Engelke K, von Andrian UH, et al. (2003) J Cell Biol 160: 267-77.

Wynn ML, Consul N, Merajver SD, Schnell S (2012) Integr Biol (Camb) 4: 1323-37.

Xavier P, Belo L, Beires J, Rebelo I, Martinez-de Oliveira J, et al. (2006) Arch Gynecol Obstet 273: 227-231.

Xia W, Liu L, Ho P, Spector NL (2004) Oncogene 23: 646-653.

Xu P, Derynck R (2010) Mol Cell 37: 551-566.

Xu P, Liu J, Sakaki-Yumoto M, Derynck R (2012) Sci Signal. 5: ra34.

Yamamoto S, Tsuda H, Takano M, Kita T, Kudoh K, et al. (2007) Mod Pathol 21: 115-124.

Yamao M, Naoki H, Ishii S (2011) PLoS One 6: e27950.

Yoshida S, Harada T, Mitsunari M, Iwabe T, Sakamoto Y, et al. (2004) J Clin Endocrinol Metab 89: 823-32.

Yoshino O, Osuga Y, Hirota Y, Koga K, Hirata T, et al. (2004) Am J Reprod Immunol 52: 306-311.

Yoshino O, Osuga Y, Koga K, Hirota Y, Hirata T, et al. (2006) J Reprod Immunol 72: 85-93.

Yousef MHM, Chai DC, Mwenda JM, D'Hooghe TM (2009) Fertil Steril 92: S11.

Yu J, Smith VA, Wang PP, Hartemink AJ, Jarvis ED (2004) Bioinformatics 20: 3594-3603.

Zaman MH, Trapani LM, Sieminski AL, Mackellar D, Gong H, et al. (2006) Proc Natl Acad Sci U S A 103: 10889-94.

Zeitvogel A, Baumann R, Starzinski-Powitz A. (2001) Am J Pathol 159: 1839-52.

Zheng B, Roach LS, Ismagilov RF (2003) J Am Chem Soc 125: 11170-11171.

Zhou WD, Yang HM, Wang Q, Su DY, Liu FA, et al. (2010) Hum Reprod 25: 3110-3116.

Zhu S, Wurdak H, Wang Y, Galkin A, Tao H, et al. (2009) Proc Natl Acad Sci U S A 106: 17025-30.

Zucker S, Cao J (2009) Cancer Biol Ther 8: 2371-2373.