Fundamentals of a Systems Biology Approach to *In Vitro* Tissue Growth

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

Richard Joseph Beck

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Robert T. Tranquillo, Advisor

May 2013

Acknowledgements

I thank my advisor, Professor Bob Tranquillo (BME, CEMS), for providing the opportunity to work in the fields of tissue engineering and proteomics. I thank committee members, Professors Dennis Cook (STAT), Tim Griffin (BMBB), Dave Odde (BME), and Wei-Shou Hu (CEMS) for their time and input.

In the Tranquillo lab: I thank Zeeshan Syedain for studies on adaptation that provided some impetus on the need for this project and for cell culture training. I thank Justin Weinbaum for training me in my first wet lab techniques such as pipetting, Westerns, cell culture, and many others. I thank Sandy Johnson for skilled technical work in developing the LOX assay and implementing the PIP ELISA. There are a number of undergraduate researchers who contributed to my work as well. I thank Lee Meier for applying his characteristic hard work and enthusiasm when assisting me with the intact phosphoprotein MS sample prep and analysis during his undergraduate research. I thank Alex Weston for thorough work on the LOX assay. I thank Chris Cheng for work with qRT-PCR, sample prep for peptide mass spec, and the LOX assay. I thank Kaitlin Medal for data analysis with MATLAB. I thank Michael W. Nelson for getting qRT-PCR up-and-running in this project. I thank Stephanie Birkholz for assisting in starting up the LOX assay. I thank Rakibul Hasan for data analysis. My thanks to all other lab members who contributed their thoughts

I thank Dennis Cook's students Zhihua (Sophia) Su and Xin (Henry) Zhang for extensive statistical analysis with challenging data. I thank Dennis, Zhihua, and Xin for explaining the stats to me.

I thank the Griffin lab for space to do peptide sample prep and discussing my questions and ideas about proteomics. I thank Tim for his mentor role and for the welcoming atmosphere of his lab. I thank Matt Stone for training me during the summer of 2009 in using Gygi's method for phosphopeptide enrichment. I thank all the staff at the Center for Mass Spectrometry and Proteomics at the University of Minnesota for their hard work and discussions on MS-based proteomics, especially Matt Stone (Orbitrap XL) and LeeAnn Higgins (Orbitrap Velos).

I thank Sue Van Riper (BICB) for developing PIN and data handling and custom scripts to integrate PIN output with MaxQuant output.

I am grateful for the resources of the Minnesota Supercomputing Institute. I thank especially Pratik Jagtap for introducing me to MaxQuant and guidance with identification and label-free quantification tools.

I would also like to acknowledge my funding resources NIH R21 EB11561 (EB=NIBIB) and

NIH 2T32 GM008347-21A1/P7BTWJ.

Abstract

Tissue engineering needs a paradigm shift in order to generate clinically useful products. The field has yet to regularly produce implantable tissue-engineered products. The conventional manner in which input stimuli are applied without consideration of current cellular activity level is certainly suboptimal.

The objective of this line of research is to produce a method for rationally choosing input stimuli that drive the cells toward optimal tissue growth. Transient phosphorylation of signaling proteins after a perturbation in stimuli contains biological information concerning downstream tissue growth. The overall project aims to build a statistical model predictive of tissue growth via information of the upstream phosphoproteome minutes after a change in stimuli. The validity of such a statistical model can be tested based on its utility to direct tissue growth: stimuli will be chosen on the basis of which corresponding phosphoproteome profile(s) is predicted to yield the best downstream tissue growth; this can be directly compared to conventional tissue engineering methods.

This doctoral project focused on obtaining sample types and tailoring methods appropriate for a systems biology and statistical approach, especially in regard to the label-free quantification of phosphopeptide enrichments.

Neonatal human dermal fibroblasts (nhDF) were expanded to near confluence, at which point basal medium for tissue production was applied. After two days, nhDF were perturbed with basal medium supplemented with 1 or 10 ng/mL TGF-β1. Cells were harvested at 10, 20, or 30 minutes for intracellular proteins. Resultant protein lysates were digested to peptides via trypsin and enriched for phosphopeptides via Iron Immobilized Metal Affinity Chromatography (IMAC). Phosphopeptide enrichments were analyzed by tandem mass spectrometry. A total of 1689 peptides were both identified with phosphorylation and quantified using distinct algorithms. Under strict statistical tests, 22 of these peptides were found to differ between

treatments/time. Corresponding downstream collagen deposition was also found to differ between treatments.

These results indicate that the type of quantitative data needed for the overall project can be acquired. The methods developed can be used in finding a statistical relationship between tissue growth and upstream phosphoproteome profiles.

Table of Contents

List of Tables

List of Figures

List of Equations

Abbreviations used

ACN – acetonitrile

BCA – bicinchoninic acid [in assay]

CID – collision-induced dissociation, also known as ion trap collision-activated dissociation

DMEM – Dulbecco's modified Eagle medium

ECM – extracellular matrix

EGF – epidermal growth factor

FBS – fetal bovine serum, also known as fetal calf serum

HCD – higher-energy collision-activated dissociation, also known as beam-type collision-activated dissociation

IMAC – immobilized metal affinity chromatography

LC – liquid chromatography

LC-MSMS – liquid chromatography tandem mass spectrometry

MALDI-TOF – matrix-assisted laser-desorption ionization time-of-flight [mass spectrometry]

MS – mass spectrometry

MSMS – tandem mass spectrometry, also known as MS^2

m/z – mass-to-charge ratio

nhDF – neonatal human dermal fibroblast

PBS – phosphate buffered saline

PIN – proximity-based intensity normalization (of mass spectrometry data)

PIP – procollagen type I C-peptide

PTM – post-translational modification

qRT-PCR – quantitative reverse transcription polymerase chain reaction

SIMAC – sequential elution from IMAC

SRM – selective reaction monitoring, also known as multiple reaction monitoring (MRM)

TFA – trifluoroacetic acid

TGF-β1 – transforming growth factor β1

Chapter 1 Introduction

Insufficiency of Conventional Tissue Engineering

The field of tissue engineering aims to deliver biological replacements to treat diseased or damaged tissues. A major pursuit of tissue engineering is the *in vitro* growth of soft connective tissue replacements, such as skin, vessels, heart valves, myocardium, ligaments, tendons, and cartilage. However, engineered tissues have failed to be implanted on any regular basis. Despite a large and increasing research effort aimed at growing functional connective tissues *in vitro*, including experimentation with a wide array of cell sources, tissue scaffolds and applied stimulation (chemical, mechanical, and/or electrical), with few exceptions they lack the required functionality. Generally this is because the tissue growth is limited and the resulting material/mechanical properties are inadequate for *in vivo* applications.

This failure is largely due to limited growth and lack of adequate mechanical integrity of *in vitro* tissues. The extent of tissue growth and strength are determined by the cells actively proliferating, synthesizing and modifying structural components.

Conventional tissue engineering begins with cells, scaffold, and input stimuli in an initial state. Then input stimuli are applied to the system at constant levels for the remainder of the culture period. This is obvious in the case of growth factors that are kept at set concentrations. Chemical, electrical, and mechanical stimulation are often applied in a similar fashion. Cyclic stretch, though dynamic, is often applied with set parameters and therefore is also a fixed input.

The long-term response is defined after weeks of culture when the mechanical and biological properties of the tissue-engineered construct are evaluated.

Many research groups, including the Tranquillo laboratory, are pursuing the paradigm of improving tissue growth by supplementing the culture medium with serum and/or specific growth factors and applying cyclic mechanical stretching in order to stimulate cell proliferation and extracellular matrix (ECM) production (Balestrini and Billiar, 2006; Engelmayr et al., 2005; Hahn et al., 2006; Isenberg and Tranquillo, 2003; Jeong et al., 2005; Niklason, 1999; Seliktar et al., 2000). While chemical and mechanical signals, alone or in combination, have led to improved tissue growth as measured by material/mechanical properties, the current paradigm of applying these signals at set levels for the entire period of tissue growth is almost certainly

suboptimal. The ubiquitous occurrence of adaptation by cells to their extracellular environment is likely a contributing factor (i.e. when a constant magnitude of cyclic stretching is applied, the initial stimulation of the stretching fades and the continued stretching no longer induces as much cell response).

Adaptation: Evidence and Mitigation

Often when cells in culture are presented a stimulus known to elicit tissue formation, initially they actively respond, but then adapt over time to basal activation. The existing tissue culture approach of maintaining the same level of cell stimulation throughout the entire culture period and resultant adaptation likely contributes to this lack of success in tissue engineering. With continual application of fixed input, cells revert to baseline activity (Turner, 1998) resulting in poor tissue growth (Syedain et al., 2008). Tissue growth likely occurs mostly during the time period immediately following the application of the new input.

Specific changes in input during the culture period have been shown to improve functional characteristics of *in vitro* tissue. For example, incremental cyclic distension significantly improved the mechanics of tubular constructs in the Tranquillo laboratory, even more than fixed cyclic stretch groups. Tubular tissue constructs were prepared from human dermal fibroblasts entrapped in fibrin gel formed in a tubular mold. The cells were allowed to remodel the fibrin for two weeks in static culture conditions. For the next three weeks they were either stretched cyclically (0.25 seconds stretched and 1.75 sec relaxed during every 2 sec interval) with 5%, 10%, or 15% strain amplitude – the conventional approach. In a fourth group, the constructs were stretched with a strain amplitude that was incremented every four days from 5% to 15% over the three weeks. As can be seen in Figure 1.1, while there was a modest (up to 50%) increase in the ultimate tensile strength (UTS) of the constructs using the conventional approach, the UTS increased almost 200% in the case where the magnitude of stretch was simply incrementally increased over the same 5 – 15% range (IS15) during the same period (Syedain et al., 2008). While this does not prove adaptation limits the conventional approach, these data are consistent with the adaptation hypothesis, and at the very least demonstrate that the conventional approach can be highly suboptimal.

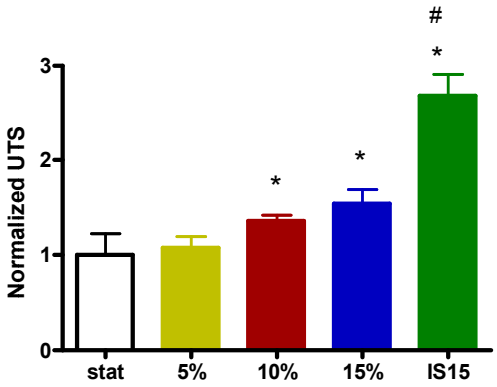

Figure 1.1: Fibrin-based tubular tissue constructs were subjected to cyclic stretching for 3 wk. An increase in strength relative to static controls is obtained by simply incrementing the strain amplitude (IS15) vs. using a constant strain amplitude.

Interestingly, in a preliminary experiment conducted using a non-monotonic strain amplitude sequence, but again with time-averaged strain of 10% as in IS15, there was no benefit of the strain amplitude step changes (Figure 1.2). This underscores the point that it is not possible to know the optimal strain amplitude regimen a priori, but one evidently exists, and the proposed approach will enable its approximation during the construct incubation.

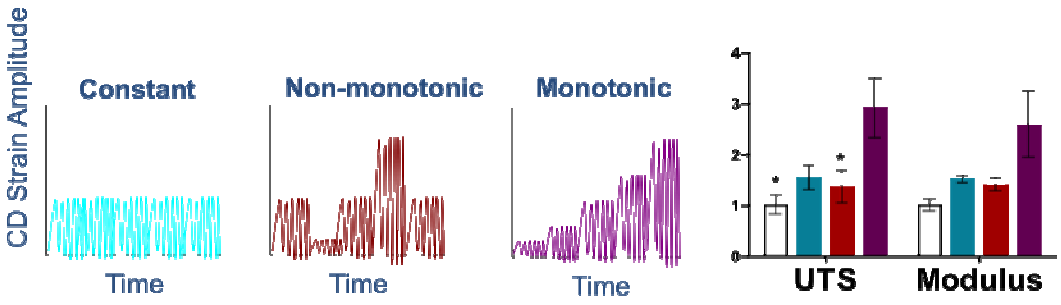

Figure 1.2: Various stretching regimens and the resultant mechanical properties normalized to static controls. (Colors in the rightmost frame correspond with the illustrated regimen. For both UTS and Modulus, first column is static control, the second column is constant 10% amplitude stretch, the third column is non-monotonic strain amplitude sequence with time-averaged strain of 10%, and the fourth column is monotonically increasing strain amplitude with time-averaged strain of 10%)

These results suggest that varying the inputs during culture can improve the tissue growth process. However, the choice of how to vary stimuli is based on intuition and lacks rationale. Tissue growth *in vivo* is inherently a complex coordinated process. This must be considered

rationally in order to recapitulate tissue growth *in vitro*. *In vivo* tissue growth during development and wound healing is driven by chemical gradients in space and time, involving sets of genes. A rationale way to vary inputs such that cells are maintained in a highly synthetic phenotype would be a large improvement over the conventional tissue engineering approach. In order to accomplish this, information about what is going on inside the cells could be applied to the tissue growth process.

Response ~ f(Input)?

Often incorporated into the current paradigm is also the attempt to (intuitively or empirically) relate input stimuli to output tissue growth response. This concept can be expressed as attempting a statistical model of using input stimuli to fully describe the response (Equation 1.1). This is fundamentally flawed, as information solely from controlled inputs cannot account for the history of uncontrolled influences on cell behavior and the current state of the cell. The same stimuli applied to a cell in one state versus another state can elicit divergent responses. The complexity of the living cell has made efforts to consistently relate inputs to outputs futile, partly because the two measures are separated by a long route of biological machinery. Finding responses that are functions of the input may be too complicated. The number and intricacy of cell signaling pathways is growing with each discovery.

$$R \sim f(I)$$

Equation 1.1 Conventional tissue engineering approach to describe the tissue growth response only with input stimuli. Response R and input I.

A Systems Biology Approach to Tissue Engineering

Cue-signal-response

Many processes in cell biology can be characterized by cue-signal-response (Janes et al., 2004a). When an extracellular input (cue) such as a growth factor is presented to cells, it binds specifically to receptor(s) on cell surfaces. The receptor then changes conformation as a result of this binding event, causing physiochemical changes on the intracellular domain of the receptor. Similarly, when mechanical stretch (cue) is applied to cells, integrin receptors (which

are bound to extracellular matrix proteins) undergo changes on the intracellular domain. Intracellular proteins come into contact with the changed receptor and they themselves are modified, both physically and functionally. These events, known as intracellular signaling, propagate the information to the nucleus, where a long-term cellular response can be formed, namely the expression of mRNAs and subsequent translation of proteins. This long term response can involve the production and assembly of proteins that are the components of tissue.

Systems Biology Approach

Thus intracellular signaling is causally related biologically to the mRNA and protein that is produced later. It is this biological causality (upstream signals cause downstream tissue growth) that we aim to describe with a predictive statistical model. Therefore it is hypothesized that data of intracellular signaling is able to be statistically related to data of the downstream response. Once a statistical model describes these upstream events to these downstream events, one need only measure the intracellular signals and use the model to predict the tissue growth response.

Intracellular signaling proteins transmit stimuli information to mechanisms that are responsible for long-term response and therefore are an intermediate between inputs and outputs. Additionally, the profile of signaling proteins inherently contains information on the current state of the cell. The information obtained from profiling signaling proteins (signals) and the mechanistic link between signals and cell outputs suggests a great potential for a statistical relationship between these signals and long-term response. Janes and colleagues (Janes et al., 2004a) used this approach with two inputs, one that leads to cell death and one that leads to cell survival. Using partial least squares regression and principal component analysis, they found a relationship between relevant signaling proteins and cell viability. The construction and implementation of a similar statistical model (Equation 1.2) in the tissue engineering context could provide a means to rationally vary input stimuli in a manner that optimizes tissue growth.

$$R \sim f(S)$$

Equation 1.2 Proposed tissue engineering approach to describe the tissue growth response with intracellular signals. Response R and signals S.

Overall Project Objective

Tissue engineering needs a paradigm shift in order to generate clinically useful tissue-engineered products. Though some researchers in this field have attempted to dynamically vary stimuli, the choice of how to vary them is based on intuition and lacks rationale. The objective of this research is to produce a method for rationally selecting input stimuli that drive the cells toward optimal tissue growth. First we will seek to build a statistical relationship between intracellular signals induced by stimuli and the downstream tissue production (Equation 1.2). Then the statistical relationship determined can be used predictably to increase production of tissue; stimuli will be chosen on the basis of said statistical relationship between signaling proteins and downstream tissue production. In other words, the input stimuli are varied such that cellular activity is maintained in a synthetic phenotype (cells are to be kept actively producing tissue, not reverting to baseline levels of activity).

The overall project aims to develop a statistical model wherein "upstream" intracellular signals (that are generated due to extracellular stimuli) predict "downstream" tissue growth. This systems biology approach considers not only the role of individual components, but also their complex interactions. The utility of this statistical relationship will be demonstrated by predicting means to increase tissue production during culture. Whereas the conventional approach relies on Response ~ f(Input) and ignores current cell state, the systems biology approach is built using information-rich data from the cell.

Assuming the adaptation hypothesis is valid, at least for this system, the challenge is then how to best apply the chemical and mechanical "input" so as to keep the cells in a synthetic phenotype and thereby maximize the response or "output". The response is fiber production of collagen and elastin, the main non-cellular structural components of soft tissue and activity of lysyl oxidase, an enzyme that cross-links these fibers. (These metrics include cell proliferation and the ECM deposition per cell.)

Hypothesis

The hypothesis of this research is that transient signaling proteins created by the cell upon stimulation are related to the downstream tissue development. More specifically, the phosphorylated protein subset of the proteome is expected to contain relevant information of short-term cellular events after stimuli. This information is thus expected to be statistically

related to measures of subsequent tissue growth. Moreover, such a statistical relationship will have the predictive capability to be used in optimizing tissue growth.

Rationale

The success of this project will be a springboard for future research, providing the tools to optimize tissue growth in three-dimensional tissue constructs to the point of functional, implantable tissue replacements. Tissue engineers will be able to periodically update the culture conditions to drive the cells away from undesired behavior (such as adaption) to desired behavior (such as tissue growth). Researchers will be able to keep cells actively synthesizing extracellular matrix. This will allow culture conditions to be optimized for maximizing mechanical strength of the tissue constructs.

Significance

This would be a paradigm shift in the field of tissue engineering.

The proposed approach to growing a tissue *in vitro*, which employs systems biology-based statistical analysis and process optimization, has the potential to significantly improve the quantity and quality of tissue that can be grown. It also offers several other potential benefits.

1) While the method will be validated using two particular controlled inputs (the concentration of a growth factor and mechanical stretching), the method could be easily translated to bioreactors using any number of controlled inputs (electrical current, perfusion-based transport of oxygen and/or glucose, etc.) that are appropriate or relevant to the application.

2) Although the application immediately proposed is for improving the mechanical properties of tissue grown in vitro, other applications that involve control of cell fate or behavior can be readily envisioned, such as improved control of stem cell differentiation by developing a signal-response relation that relates intracellular signals to differentiated phenotype and then using the relation to guide the culture conditions during differentiation.

3) While the signal-response relation itself is empirical, it serves to identify intracellular signals that are correlated to cell behaviors of interest and can lead to discovery of new

key signals or groups of signals that will stimulate and guide research aimed at elucidating signaling networks.

Research Design

The vision for the overall project is to develop a rational basis for optimizing *in vitro* tissue growth via combined periodic changes in stimulation conditions that are predicted to maximize a response of interest, wherein the "downstream" cell response R of interest are related to "upstream" intracellular signals S (that are generated due to extracellular stimuli). This will be accomplished via a signal-response statistical model without requiring any knowledge of signaling networks leading to the ultimate cell response of interest.

The success of the statistical approach relies on generating the maximum amount of accurate, quantitative information regarding intracellular signals. The signaling proteins which span cell signaling space are likely from multiple signaling pathways. Thus, a mass spectrometry-based proteomics approach is used to capture changes in protein phosphorylation correlated with improved tissue growth.

Protein phosphorylation is chosen as the biochemical signal based on its role in cellular signaling, and the expectation that it will be among the fastest and most direct cellular events after perturbations in stimuli. In fact, protein phosphorylation has been shown to be a key factor occurring on short time scales in the collagen deposition pathway (including Smads (Verrecchia and Mauviel, 2002)) and in cell cycle regulation (such as Rb (Guo, 2005)).

Note: This doctoral project involved only Aim 1. This aim was not completed and Aims 2 and 3 were not begun as they depend on the completion of Aim 1.

Specific Aim 1

Determine the statistical relationship, R~f(S) (Equation 1.2), between intracellular signals induced by stimuli and the deposition of collagen and elastin for cells in two-dimensional culture.

For this systems biology approach to tissue engineering, first a relationship between long-term cellular response, R and short-term intracellular signals, S must be determined. Intracellular signaling, S, contains information on how the cell is responding a short time after a perturbation. Over time, the cell integrates continual information from signaling proteins in order to respond. This long-term response, R, involves the formation, accumulation, and organization of extracellular matrix (in particular collagen and elastin), cell proliferation, and other tissue growth characteristics that can be measured. In order to 'tease out' what information in the signals, S, contribute to a response, R, it is necessary to compare responses from different signaling histories. The time points for measuring signals will be determined in Aim 1 as the shortest time points that yield the best statistical relationship in terms of maximum capture of the ultimate collagen and elastin deposition data.

The statistical description will be developed by first exposing cells to a range of stimulation conditions, quantifying the phosphoproteome (which is expected to include a sufficient number of relevant upstream intracellular signals, such as Smad proteins in the collagen expression pathway), and measuring the cell response R of interest. These data will be analyzed for the aforementioned signal-response statistical relation. Recently developed statistical theory and methodology has potential for this application (Cook and Ni, 2005, 2006; Cook, 2007). The choice of statistical methodology depends on the nature of the measured data. Statistical models using proteomics data are being accomplished (Janes et al., 2004b; Kumar et al., 2007; Miller-Jensen et al., 2007). The number of experiments required to build a satisfactory statistical model is dependent on the nature of the data and the type of statistical analysis chosen.

It is likely that the conditions that, for example, result in maximum collagen deposition will differ from those that result in maximum elastin deposition. In that case optimization of their combined deposition will be made.

Specific Aim 2

Demonstrate that the statistical relationship determined in Aim 1 can be used predictively to increase production of collagen and elastin for cells in two-dimensional culture.

Once a relationship between signals and tissue growth is established, its utility will be the optimization of tissue growth. It will identify a hopefully low-dimensional function f(S) of intracellular signals S that carries all the relevant available information about the long-term cell response R. It will be used in the same system to rationally choose the current stimuli conditions. Having determined such a signal-response relation, we can then proceed to optimizing tissue growth using an iterative optimization strategy. The relationship between signals and response will be used to direct the cells toward the best predicted response.

This will require periodic interrogations of the cells by various perturbations to elicit corresponding S (Figure 1.3). The statistical model relating signals to response will be performed on the data from signaling protein analysis to find the signal profile which elicits the best predicted response, that which has, for example, the maximum amount of collagen and elastin production. The input associated with the generated the signals for the best response will be implemented until the next interrogation. At the time of the next interrogation, the state of the signaling proteins has changed due to adaptation or the effects of uncontrolled stimuli (e.g. collagen that has been deposited). Interrogations will be iterated to continue stimulation throughout the culture period until harvest. Thus input stimuli are rationally varied in this method.

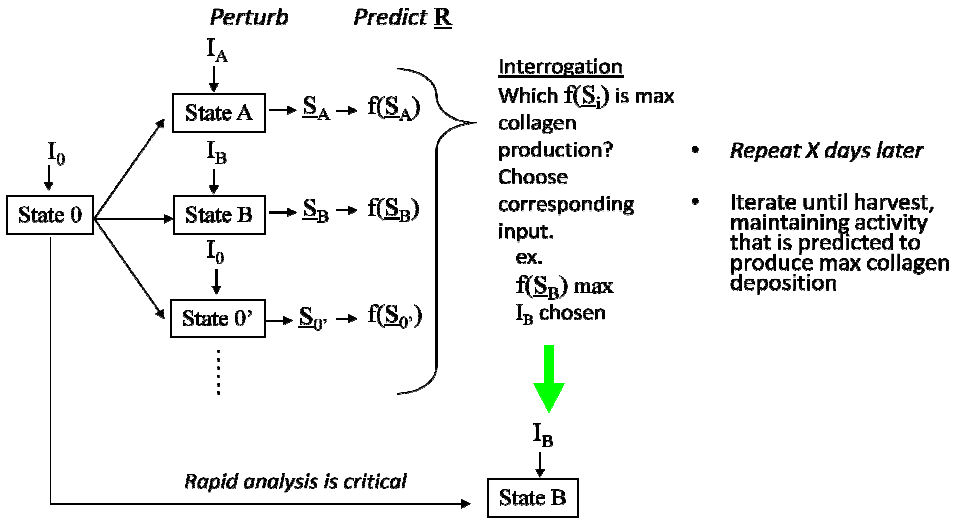

Figure 1.3: Once a model R~f(S) is described, it can be used to drive the cells toward optimal tissue growth.

The following is a more detailed explanation of the approach: Cells in culture of the same cell type and with the same system used in developing the statistical relationship will be stimulated initially with conditions known to enhance tissue growth, such as soluble TGF-β1 and/or mechanical stretching. After X days, the stimulus will be changed to prevent adaptation via an interrogation process of short-term step-response experiments. A portion of the cells will be divided into groups for parallel perturbations and subsequent harvest. (The remaining cells which do not experience perturbation at this time will be used to assess downstream tissue deposition or future interrogations.) For example, in one group the TGF-β1 concentration will be increased, in other groups the concentration will be increased by different amounts, and in a control group the concentration will remain the same. Within 30 min after this perturbation in stimulus (and possibly at multiple time points), each group will be harvested for their phosphoproteins/phosphopeptides. These will be analyzed by mass spectrometry as described in Chapter 2 Signals Measurement. The resulting data, S, will then be entered into the predetermined statistical function, f(S), to predict long-term collagen and elastin response (R) for each perturbation group. The phosphoprotein/phosphopeptide profile that corresponds to the downstream response of the most collagen and elastin production will be elicited in all remaining cells by applying the corresponding input stimulus. It is essential that the update in stimulus conditions be made as soon after the interrogation as possible as the current cell state will change due to adaption; our goal is within one day. This stimulus will continue for the next X days, when another interrogation will be done in a similar manner as before to elicit phosphoprotein/phosphopeptide profiles that will be used to update the stimulus. The process will continue repeating until the final harvest date when the remaining cells will be analyzed for extent of tissue growth. Control groups such as fixed stimuli and stimuli dynamically varied according to intuition rather than rationale will be run in parallel the full length of the study. This optimization procedure that makes use of the statistical relationship (Aim 1) is depicted in Figure 1.4, which provides an overview of the Research Design.

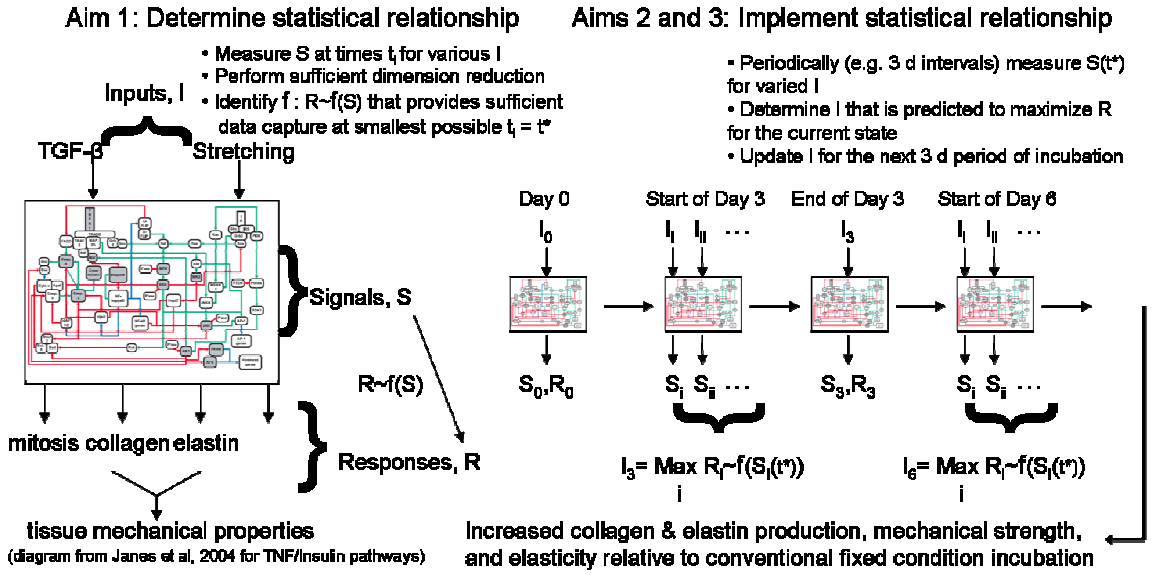

Figure 1.4: Overview of Research Design and Specific Aims.

The robustness of the statistical relationship will first be validated by using a stimulus regimen that was not used in determining the statistical relationship (e.g. incremental strain-and-hold). Robustness is expected because the success of the statistical relationship only depends on inclusion of the operative intracellular signals - it does not depend on how they are generated. Thus, complicated phenomena such as autocrine factors being released in response to cyclic stretching are not problematic as long as their effects do not involve intracellular signals not used in developing the statistical relationship. Information from signaling proteins and downstream tissue growth from this aim can be used to improve the model generated in Aim 1.

Specific Aim 3

Demonstrate that the statistical relationship determined in Aim 1 can be used predictively to increase deposition of collagen and elastin fibers for cells in a more applicable three-dimensional geometry with improved mechanical properties of the resulting tissue construct.

This Aim is analogous to Aim 2, but to be performed in the more challenging, but more relevant, context of tissue growth in a three-dimensional construct. The same cell type used in the previous aims will be used. The experimental design will be almost identical to Aim 2, except

that a three-dimensional tissue construct will result and tensile mechanical properties and histology will also be characterized.

To establish proof-of-principle, initial experiments use a convenient two-dimensional cell culture on a surface. After this proof, experiments will be in a tissue construct of geometry applicable to implantation.

Signals

Suitable signals obtained soon after a perturbation in stimuli should contain sufficient information on the current cell state to build R~f(S). The phosphoproteome, the set of all phosphorylated proteins/peptides, was the chosen candidate for the Signals measurement.

Phosphorylation

Intracellular signaling is driven primarily through changes in networks of proteins. Phosphoproteins are a particularly relevant subset of intracellular signaling proteins. Phosphoproteins are a subset of key intracellular signals that are generated due to extracellular stimuli and lead to changes in downstream processes, such as tissue growth. Phosphorylation of proteins is a major means by which information is propagated through intracellular signaling proteins. Phosphorylation, a post-translational modification (PTM), is the addition of a phosphate group to an amino acid in the protein. Serine, threonine, tyrosine, histidine, and aspartate are known targets (His and Asp phosphorylations are so labile that the phosphate falls off before MS analysis). After perturbation, changes in phosphorylation can occur within seconds to minutes, making this PTM class potentially useful for obtaining information on S within time constraints. This event is reversible; the functional regulators are proteins themselves: a kinase adds the phosphate and phosphatase removes the phosphate. Phosphorylation changes the function of its target protein, leading to activation or deactivation.

The main goals of the field of phosphoproteomics are to determine the biochemical and cellular function(s) of changes in phosphorylation. In the case of tissue engineering, we have a known overall function of tissue growth; specific biochemical an cellular functions are not the focus of this project, though their elucidation may have usefulness in tissue culture (ex. Inputs that act on a specific network or protein to encourage a specific biochemical or cellular function that contributes to the overall tissue growth).

Phosphorylation Enrichment

The addition of a phosphate group in renders the affected amino acid chemically distinct from non-phosphorylated residues. Therefore, strategies have been developed to enrich for phosphorylated proteins and peptides. The analysis of phosphorylated proteins, or phosphoproteomics, has been gaining momentum recently due to the discovery and refinement of tools that can chemically capture the phosphate group(s) of phosphoproteins or phosphopeptides.

In Immobilized Metal Affinity Chromatography (IMAC), metal such as iron (III) interacts with the negatively charged phosphate group(s), making enrichment of phosphoproteins or phosphopeptides possible. Titanium dioxide (TiO_2) enriches for phosphopeptides with a different chemistry than IMAC (Thingholm et al., 2006). Sequential elution from IMAC (SIMAC) following TiO_2 enrichment has been developed to separate multiply phosphorylated peptides from singularly phosphorylated peptides (Thingholm et al., 2008; Ye et al., 2010). Mono-phosphopeptides do not bind as strongly to IMAC material and thus the buffers can be optimized to either keep them in the flow through or eluted first. A stronger elution buffer is then used to obtain the multi-phosphopeptides.

Antibodies for phosphotyrosine allow specific capture of peptides containing this residue, which represents selective subset of phosphopeptides (~1% of phosphopeptides). Phosphotyrosines have been shown to contain rich information on a cell's early signaling events.

Mass Spectrometry in Proteomics

A critical tool in proteomics, the study of all proteins in the cell, is the mass spectrometer. Mass spectrometry is used to analyze intact proteins or peptides (cleaved parts of proteins). Tools from quantitative phosphoproteomics will help accomplish the overall project objective. Mass spectrometry profiles of sets of key signaling proteins contain information on current cellular activity.

Mass spectrometry measures proteins by their mass-to-charge ratio and relative abundance. Proteins come in a variety of masses, and when in certain solutions can carry charges. Different proteins have different affinity to pick up charges. Thus, the relative abundance displayed in mass spectra does not always reflect the actual relative abundance in the sample.

Limitations of Mass Spectrometry

Dynamic range is limited in mass spectrometry analysis; abundant proteins that stay relatively constant between samples from different treatment groups may mask where the actual changes might be occurring: in the less abundant proteins, which may be the most biologically relevant. A significant amount of time is necessary for complex sample preparation, data acquisition and interpretation. Mass spectrometry is also expensive when doing replication and a number of treatments and/or time points, due to an increase in instrument analysis time.

Response

For tissue growth Response, we have focused on collagen, elastin, lysyl osidase, cellularity, and total ECM protein. Engineered tissues need to be strong with appropriate elastic properties, enabling them to resist mechanical failure in loading situations relavant to *in vivo* function. Collagen is the main structural component of soft tissue and currently the response variable of most interest. Elastin, as its name implies, imparts elasticity. Lysyl oxidase (LOX) is a secreted enzyme that catalyzes final step in cross-linking collagen and elastin. Cellularity and total ECM protein are of general use.

Potential Challenges and Alternative Approaches

Time is an important factor in this approach of optimizing tissue growth. Updating with the chosen stimulus after interrogation through statistical analysis assumes that the cell is still in a state similar to when the signals were harvested at interrogation. The steepest descent strategy requires doing multiple "interrogations" to deduce the optimal inputs to maximize the outputs at each interrogation/optimization cycle. The total time to do these interrogations and make the deduction must be significantly less than the duration of the tissue growth period. In Figure 1.4, a 3-day cycle means the interrogations and deduction should not take more than one day so the bioreactor conditions can be reset for the remaining two days of the cycle. Interrogations can be done in parallel, which will significantly reduce the time required. Sample handling and phosphoenrichment steps will need to be performed efficiently. Peptide mass spectrometry was initially dismissed in preliminary work due to the typical protocol time for trypsin digestions, an overnight step. However, there are possibilities to optimize this reaction to shorter times, such as application of high pressure (Akasaka et al., 2008; López-Ferrer et al., 2008) ultrasound (Priego-Capote and de Castro, 2007), or simpler methods such as increasing the enzyme to

substrate ratio. The peptide mass spectrometry methods discussed offer better resolution and dynamic range than intact protein mass spectrometry. Mass spectrometry and MS/MS involve time consuming steps for sample preparation, data acquisition, and data interpretation. This is important to keep in mind for the aforementioned one-day time constraint as MS analysis will have to be completed in a short amount of time. During the Aim 1 stage of the research the focus must remain on achieving a statistical relationship between short-term and long-term response irrespective of the time required. The steps of mass spectrometry and MS/MS can later be optimized for time efficiency. In the proposed targeted MS/MS approach for Aims 2 and 3, new phosphoproteins that contribute to f(S) will be missed in favor of timely analysis. If a thorough MS/MS discovery analysis is possible within the time constraints for Aims 2 and 3, potential contributors could be identified.

The one-day time constraint may not be short enough. In Aims 2 and 3, if the cells adapt before the update in stimulus, the cells may not be properly activated by the chosen stimulus update. An additional sampling of signaling proteins during Aim 1 (at the projected time after interrogation that the update would be made in Aim 2) could provide evidence of such a possibility. In response to such a case, speed of sample processing could be improved. Statistical models that account for changes in cell state between the time of interrogation and the time of stimulus update could also be explored, using the additional data in Aim 1 and iterative data in Aims 2 and 3.

A critical issue is whether a reliable statistical relationship exists between sufficiently short-term signals (within the one-day constraint to interrogate, analyze, and update) and the long-term proliferation and specific (per cell) collagen/elastin deposition rates. Data from mass spectrometry of phosphopeptide enrichments alone may not contain sufficient information to explain/predict downstream tissue development. In such a case, complementary data can be extracted from some of the many technologies and methods that exist for investigating short-term cellular response to stimuli such as western blot, ELISA, and kinase assay. Differential expression has been revealed by western blotting in our laboratory. This method has been used in the construction of cell-fate models by others (Kumar et al., 2007). Mass spectrometry is more advantageous than western blotting in cases where the proteins of interest are not known a priori. However, protein species involved in relevant biochemical networks continue to be

confirmed, so probing for known proteins may provide information that would be used in place of or in addition to the mass spectrometry data.

In Aim 1 and 2, as the deposited protein will likely be insufficient to produce a tissue film that can be mechanically tested to indicate the real output of interest ("functional protein"), lysyl oxidase can be measured via ELISA (or its activity fluorometrically) as an indicator of functional protein. Collagen and/or elastin transcription could also be used as an alternative "down-stream" output. The Tranquillo Laboratory has developed luciferase-based collagen transcription reporter cells (Weinbaum et al., 2010), which could be used in these studies in lieu of the human fibroblasts. RT-PCR could also be of utility, although the sample preparation and analysis time may preclude accomplishing a complete interrogation in one day, which is our objective.

The cell experience differs when generating the statistical model in Aim 1 versus implementing the optimization strategy in Aim 2. For model generation involving a single perturbation, the cells experience fixed inputs for most of the culture period. During optimization culture, the cells undergo changes in stimuli every few days. Aim 1 will be achieved through the use of fixed applied stimuli as in the current paradigm. This will surely result in suboptimal tissue growth. The statistical relation therefore, will be between early signals and downstream suboptimal tissue deposition. It is expected that Aim 2 will be able to optimize tissue growth to a certain extent. However, the model from Aim 1 can be iteratively improved with the data obtained in Aim 2.

In Aim 3, the statistical relationship from Aim 1 may not be optimal because the three-dimensional system will provide different external cues, and thus the signaling state may differ as well. If application of the statistical relationship in Aim 1 does not maximize tissue growth, then Aim 1 will be repeated for the system of cells entrapped in a fibrin tissue construct. This is analogous to how Aim 1 and Aim 2 share the same geometry and cell type. Data obtained in Aim 3 may be used to improve the statistical relationship iteratively.

Chapter 2 Signals Measurement

Intact Protein Mass Spectrometry

Background

Intact proteins are challenging to analyze due to their large size (on a molecular scale), complex structure and chemical properties. Analysis of intact proteins with mass spectrometry often utilizes a Matrix-Assisted-Laser Desorption Ionization Time-of-Flight spectrometer. Matrix-Assisted-Laser Desorption Ionization (MALDI) takes samples from the solid phase to ions. MALDI is useful for intact proteins, as it can ionize macromolecules up to hundreds of kilodaltons in molecular weight. For this type of ionization, the sample is dried in matrix. The matrix transfers energy from the laser to ionize proteins. A matrix appropriate for intact protein ionization is sinapinic acid; because it transfers a low amount of kinetic energy, intact proteins are not likely to be broken apart. During MALDI, usually one charge is added to each protein. Sample is taken from solid phase to gas phase by a laser. The crystalline matrix which holds the proteins must be aromatic and absorb UV light. In TOF MS, particles with small m/z arrive at the detector first. As shown in the bottom panel of Figure 2.1, when a spot with matrix alone is analyzed, background signal is observed at low m/z values.

Method Development and Troubleshooting

Due to the time constraint of analysis during the optimization phase of the overall project, analyzing complex mixtures of intact proteins via MALDI-TOF MS was viewed as an expedient approach to obtain information on S. Initially intact phosphoproteins were investigated. MALDI-TOF was used to analyze eluates from a phosphotyrosine immunoprecipitation (pY IP) kit. Phosphoprotein enrichments obtained via IMAC were also analyzed. The buffers used for both of these techniques were found to be incompatible with mass spectrometry, requiring further sample preparation. Additionally, intact protein MS may not be suitable for analyzing complex mixtures.

Intact proteins were analyzed using MALDI-TOF. Samples with low signal or no sample (matrix alone) consistently have higher noise at lower m/z values (Figure 2.1)

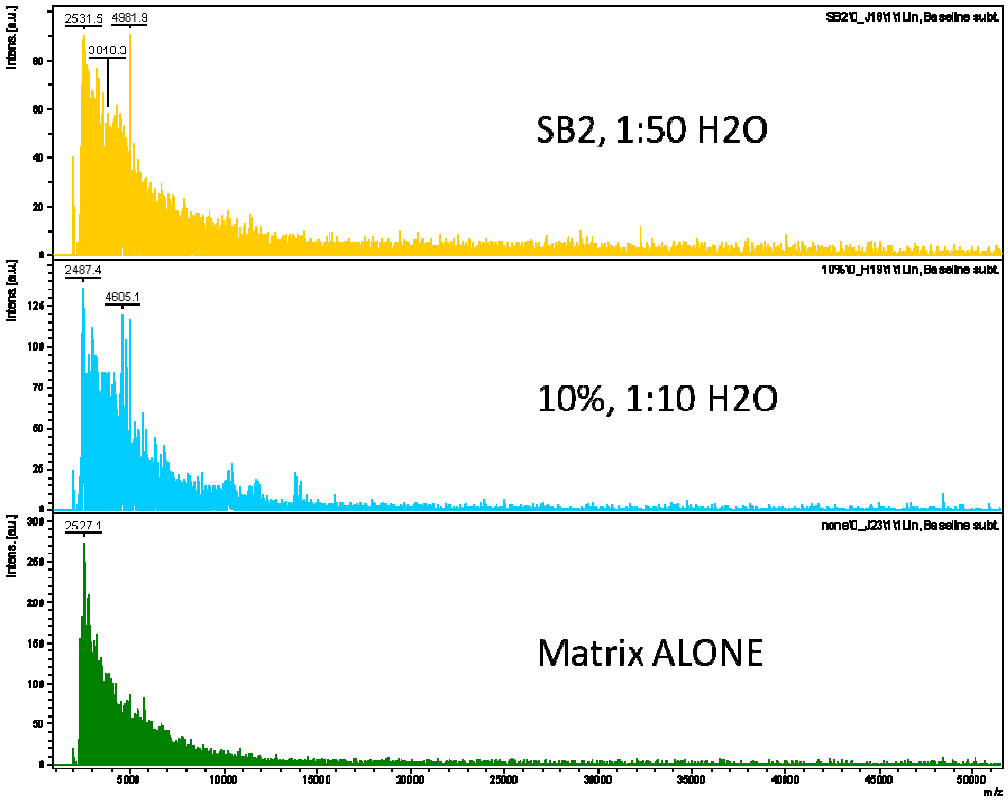

Figure 2.1 Representative spectra of low signal and matrix alone spots. Top Panel: a 1:50 dilution of a sample. Middle Panel: a 1:10 dilution of a sample. Bottom Panel: matrix alone (no sample).

Phosphotyrosine Immunoprecipitation (IP) with Intact Proteins

The immunoprecipitation kit described in Chapter 3 Methods was used to enrich for phosphotyrosine-containing proteins. Resulting eluates were analyzed via Western blot and MALDI-TOF MS. MS of these samples gave spectra with few peaks and much noise.

Early on in the project the idea of using trypsin to digest the proteins before or after IP was discussed, though the concept was not well understood by those involved with the project at the time. Peptide mass spectrometry was initially dismissed in preliminary work due to the typical protocol time for trypsin digestions, an overnight step.

Upon discussions with Millipore, it was revealed that the elution buffers are incompatible with mass spectrometry. These buffers have a salt component on the order of 1.0 M, which is too much for a ZipTip® (Millipore) to clean. Dialysis was implemented to remove the salt. Resulting spectra were of similar quality as without dialysis.

Phosphoprotein Enrichment via IMAC

Protein lysates were enriched for phosphorylated proteins using an immobilized metal affinity chromatography (IMAC) column from Pierce (as described in Chapter 3 Methods). (Samples were generated similarly to the IP for phosphotyrosine-containing proteins.)

Discussion

Over a year was spent on troubleshooting this method. Improvements included refined sample preparation such as the use of phosphoric acid (Kjellström and Jensen, 2004) and desalting, adjusting instrument settings, data analysis, literature searches, and seeking the advice of personnel at the Center for Mass Spectrometry and Proteomics at the University of Minnesota.

Though less complex than original lysates, these phosphoprotein enrichments could contain tens of thousands of various proteins. Identity and quantity has been determined for small numbers of different proteins in simple mixtures or the most abundant proteins in complex mixtures such as plasma or other clinically useful fluids (Zhang et al., 2006). With this level of complexity, identification and quantification of individual proteins of low abundance is not feasible with current technologies.

Phosphopeptide Tandem Mass Spectrometry

Background

Peptide tandem mass spectrometry (MSMS) has several advantages over intact protein MS for analyzing the phosphoproteome. MSMS can be used for identification and quantification of analytes in complex samples. Peptides are linear parts of proteins often generated by enzymatic digestion. They are structurally simpler than intact proteins, thus more amenable to chromatography techniques that can simplify the complexity of the sample. MSMS is routinely used to determine the site of phosphorylation in the amino acid sequence of identified peptides.

Tandem mass spectrometry (MSMS) is the main tool for multiplexed proteomics analysis. MSMS can be performed on samples comprising of peptides. MSMS is capable of generating data from which identification and quantification of thousands of peptides can be obtained.

Sample Preparation

The main steps in sample preparation for tandem MS are enzymatic digestion, enrichment/fractionation, and sample clean up. Trypsin is commonly used for digesting intact proteins into peptides; trypsin cleaves after every lysine and arginine, which is useful for determining amino acid sequence. Simplifying the complexity of the peptide mixture increases the depth and reproducibility of MSMS analysis. Enrichment or purification obtains a subset of the proteome. Fractionation separates analytes according to some physical property. Many MSMS instruments have a LC inline with the mass spectrometer, usually separating peptides by reverse phase. Sample clean up involves removing MS-incompatible substances such as salts or detergents.

Identification

In a typical LC-MSMS workflow, tryptic peptides are loaded into the mass spectrometer in liquid phase. Peptides are eluted from a reverse phase LC column by solvent gradient; mass spectrometry analysis of these eluting peptides occurs during the time of the LC run. The sample is ionized via electrospray ionization (ESI) or another ionization method. A mass spectrometry scan is made, capturing a snapshot of the peptides ionizing during that part of the LC run. The instrument is programmed to select the most abundant peaks for further analysis. When a peak is selected, the instrument only allows a narrow m/z around the precursor ion to proceed to dissociation. A MS scan is made on the product or fragment ions. Precursor scans, dissociation, and product scans are iterated during the course of the LC run. After the run is complete, the spectrum of products ions is often compared to a database of theoretical product ion spectra to determine identification by best fit (within specified tolerances).

Dissociation

In collision-induced dissociation (CID), also known as ion trap collision-activated dissociation, helium collides with the precursor ion to induce resonance of a narrow m/z range (ideally one peptide species), the kinetic excitation slowly heats the precursor ions, and there are resulting cleavages along the backbone of the precursors, creating product ions. These products are not fragmented further because they are now different m/z values than their precursor and therefore fragment at a different frequency. For phosphopeptides undergoing CID, the preferential cleavage often occurs at the phosphate group, resulting in a neutral loss of

phosphate and the most abundant product ion is the precursor without the phosphate. Because the b and y ions are much less abundant than the dominant neutral loss peak, these spectra are not easily sequenced.

Higher-energy collision-activated dissociation (HCD), also known as beam-type collision-activated dissociation, uses elevated vibrational energy to break bonds and fragment precursors. HCD has been shown by some to be beneficial for phosphopeptide fragmentation (Zhang et al., 2009). Neutral loss peaks are less likely to dominant phosphopeptide spectra, as the neutral loss product is more likely than in CID to undergo further fragmentation.

Method Development and Troubleshooting

First attempts at using tandem mass spectrometry (LC-MSMS) phosphopeptide analysis on the LTQ-Orbitrap XL (Thermo Scientific) resulted in identifying mostly serum proteins from FBS and no phosphopeptide spectra were reliable. Nate Tedford (Laboratory of Forest White, Massachusetts Institute of Technology) recommended a rinse with serum-free medium in addition to the HEPES rinse to better remove residual serum proteins and changing the 18 hour trypsin digestion from 37 °C to room temperature. It was also possible that the IMAC beads may have been stored incorrectly; it was then decided that storing IMAC beads in 1.5 mL tubes upright at -20 °C immediately after aliquoting may be appropriate. After implementing these changes, phosphorylated peptides were successfully identified. Immobilized Metal Affinity Chromatography was used to enrich for phosphopeptides. Following mass spectrometry analysis, instrument raw data was in a label-free quantification technique using the proteomics software MaxQuant.

Phosphopeptide Enrichments

A step change of 0 or 1 ng/mL TGF-β1 was used to perturb nhDF cultures (control and treatment groups each in triplicate). Lysis was performed 15 min after this step change. Phosphopeptide enrichment by means of immobilized metal affinity chromatography (IMAC) was performed. Briefly, proteins were trypsin digested and resultant peptides were desalted and enriched for phosphopeptides via IMAC. The phosphopeptide enrichments were analyzed by LC-MSMS on a LTQ-Orbitrap XL mass spectrometer. Protein identification from the tryptic phosphopeptides was accomplished using SEQUEST®. The average number of proteins identified per sample at < 1% FDR, >95% peptide probability, and >10ppm were 289±45 phosphoproteins and 413±25 total

proteins (phosphoprotein specificity 70%±8%), according to Scaffold (v2.06). The replicates for the control group had 175 phosphoproteins in common. Similarly, the 1 ng/mL TGF-β1 replicates had 203 phosphoproteins. The number of phosphoproteins in common between all 6 samples was 130.

We continued data analysis on the experiment, using a label-free method for quantification. Each raw file from the LTQ-Orbitrap XL was processed in the Quant module of MaxQuant (v1.0.13.13; developed by Matthias Mann's group) and Mascot (Matrix Science). We employed the Identify module of MaxQuant at a stringent 1% False Discovery Rate (FDR) for peptide, protein, and site identification using the human IPI database v3.52.

Only proteins identified in every treatment and control replicate were analyzed further. Each protein's intensity was normalized by the median of intensities from the corresponding sample. A two-tailed, equal variance, Student's T-Test was performed on the scaled intensities for each protein. The fold change of the TGF-β1 group compared to the control group was determined for each protein of significantly different ($p<0.1$) scaled intensities. Each of the 26 phosphoproteins was identified with at least one phosphorylated serine or threonine. However, this procedure did not account for multiple comparisons. This is necessary to account for false positives because the T-test is being performed hundreds of times. With this consideration, none of the proteins were significantly different between control and treatment scaled intensities. This result spurred further analysis of phosphoproteins as upstream signals, including information at peptide and phosphosite levels.

Phosphotyrosine Immunoprecipitation for Peptide MSMS

Initially, phosphotyrosine immunoprecipitation was performed without further IMAC for a simpler procedure. Forest White's protocol uses 2 anti-pY antibodies: P-Tyr-100 (Cell Signaling Technologies) and PT66 (Sigma Aldrich). The White protocol was compared to one using 4G10 Platinum (Millipore), a cocktail of anti-phosphotyrosines 4G10 and PY20 (Figure 2.2). The protocols had a 39% overlap in identifications of phosphotyrosine-containing proteins, but only 14 phosphotyrosine sites were identified in total. Meanwhile, the IMAC phosphopeptide enrichment was improved in quality, such that phosphotyrosines were identified as often in phosphopeptide enrichment as phosphotyrosine immunoprecipitation.

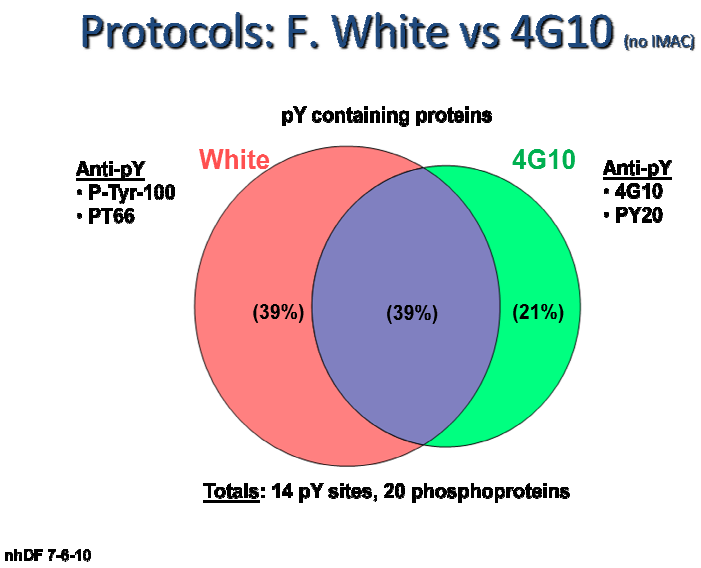

Figure 2.2: Comparison of the Forest White laboratory protocol and a protocol using 4G10-Platinum. These 2 protocols were run in duplicate.

Data Analysis

Rigged

In order to validate the accuracy of the label-free quantification workflow, a biological sample was divided into qualitative similar aliquots and loaded onto the Orbitrap Velos (Thermo Scientific) in predetermined ratios. Tryptic peptides pooled from cell lysates (nhDF 6-30-11 (0.1% vs. 10% FBS)) were divided into six 100 µg aliquots and enriched for phosphopeptides using IMAC (Figure 2.3). The resulting eluates were mixed, then the equivalent of one elution (estimated at ~ 1 µg) was aliquoted in duplicate ('Standard'), the equivalent of 1/5th of an elution (estimated at ~ 0.2 µg) was aliquoted in duplicate ('1/5th'), and 0.8 µg of peptides from the IMAC flow through (mostly non-phosphopeptides) was mixed with the equivalent of 1/5th of an elution (estimated at ~ 0.2 µg) was aliquoted in duplicate ('1/5th + 4/5th'). The '1/5th + 4/5th' samples were included to have an equal loading as 'Standard' with ~1/5th the phosphopeptides, as comparisons between unequal loadings (such as 'Standard' versus '1/5th') can be problematic.

Figure 2.3: Analytical design testing the validity of quantification methods of phosphopeptide enrichments.

The expected results therefore would be that the phosphopeptides in 'Standard' have a 5-fold higher intensity than '1/5th' and/or '1/5th + 4/5th'. By using the normalized peptide signals intensities provided by proximity-based intensity normalization (PIN), the mean fold change was 6.16 and the median fold change was 5.20. Using non-normalized intensities yielded mean fold changes of 4.51 and 3.27, median fold changes of 2.60 and 2.72 at the phosphopeptide and phosphoprotein level, respectively. The non-normalized intensities have more skew in their distributions that when using PIN (Figure 2.4).

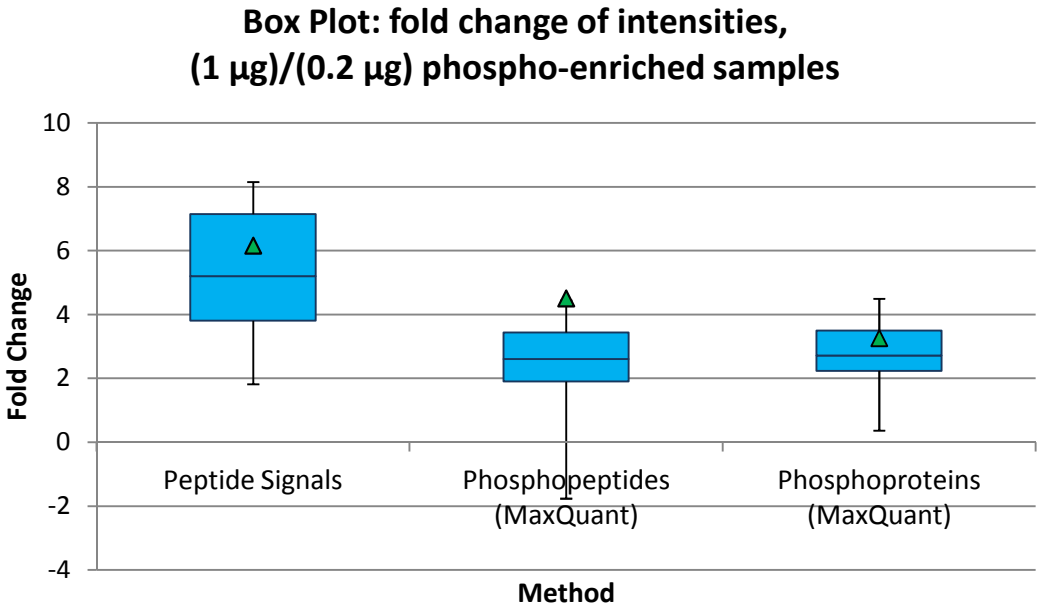

Figure 2.4: PIN peptide signals have improved accuracy when compared to MaxQuant quantification.

Chapter 3 Methods

Brief

Signals, S. Protein phosphorylation, a post-translational modification, is a major mechanism of intracellular signaling. Cells are lysed in an 8 M urea buffer and proteins are digested to peptides with trypsin. Immobilized affinity chromatography (IMAC) is performed on the peptides to enrich for phosphopeptides. Phosphopeptides are then analyzed by LC-MSMS in a LTQ Orbitrap Velos (Thermo Scientific) mass spectrometer.

Response, R. Downstream events that capture the dynamics of peak and basal tissue growth could be used to determine the area under the curve of cellular activity. RNA is extracted from lysed cells; cDNA is synthesized from the RNA using reverse transcriptase. The resulting cDNA is analyzed via quantitative real time PCR. Soluble collagen is analyzed with an ELISA method. Activity of secreted lysyl oxidase (LOX), an enzyme that crosslinks collagen, is analyzed fluorometrically. Insoluble collagen fibers are quantified with a hydroxyproline assay.

Cell culture

Expansion

Cells of type neonatal human dermal fibroblast (nhDF), passages 7 – 9 were used in the experiments described. Dermal fibroblasts (skin cells) were suspended in expansion medium, 50% DMEM 50% F12 medium (Life Technologies) that was supplemented with 15% fetal bovine serum (FBS), 100 U/ml penicillin and 100 µg/ml streptomycin. Volumes were applied so that the cell density was 8,500 cells/cm^2 (based on the growth area of the plate/flask). The cell suspension was kept well mixed during this plating procedure.

Tissue growth

Cells were grown to near confluence, then switched to DMEM medium (Life Technologies) with 10% or 1% FBS, 100 U/ml penicillin and 100 µg/ml streptomycin. In some cases, DMEM without phenol red and/or sodium pyruvate was used. Some experiments included 2 µg/ml insulin and 50 µg/ml ascorbic acid to mimic culture conditions of tissue constructs grown in the Tranquillo lab. Perturbations were made using transforming growth factor β1 (TGF-β1 from R&D Systems).

General Cell Assays

Bicinchoninic Acid Assay for Total Protein Concentration

Protein concentration was determined by assay with bicinchoninic acid (Thermo Scientific Pierce). Briefly, bovine serum albumin was used for standards in a range of 0 – 200 µg/mL in increments of 25 µg/mL and diluted with 1/20th lysis buffer or PBS. Samples were diluted 1:20 in water or PBS.

Western Blot

To validate presence of phosphorylated proteins, Western blots were performed. The 4G10 (Millipore) pan-specific antibody for phosphotyrosine, ERK and phosphorylated ERK antibodies (Cell Signaling Technologies), and β-actin were used as primary antibodies. A secondary horse radish peroxidase conjugated was used.

Mass Spectrometry in Proteomics

Intact Phosphoprotein Analysis

Cell lysis

Initially used lysis buffer containing 50 mM Tris pH 7.4, 450 mM NaCl, 50 mM sodium fluoride, 10% v/v glycerol, 1% v/v Igepal/NP-40, and 2 mM EDTA pH 8.0, supplemented freshly with sodium deoxycholate and inhibitors aprotinin, pepstatin, and leupeptin.

The harvested cells were rinsed to remove metabolites and were lysed by chemical (detergents) and mechanical (sonic dismembranation) means. The buffer for lysis contains inhibitors to enzymes that degrade proteins of interest within the cell. The cellular protein fraction was separated from other cellular debris by bench top centrifugation.

Protein Phosphotyrosine Immunoprecipitation

The Catch and Release® Phosphotyrosine, clone 4G10® kit (Millipore) was used according to manufacturer's specifications using either denaturing or non-denaturing elution buffers. Note that the elution buffers for this product, as described in Chapter 2 Signals Measurement, are not compatible with mass spectrometry analysis.

IMAC Phosphoprotein Enrichment

The Thermo Scientific Pierce Phosphoprotein Enrichment Kit (work flow outlined in Figure 3.1) was tested by Pierce with a minimum of 1.0 mg of protein in lysate. This would be two 6-well plates per sample.

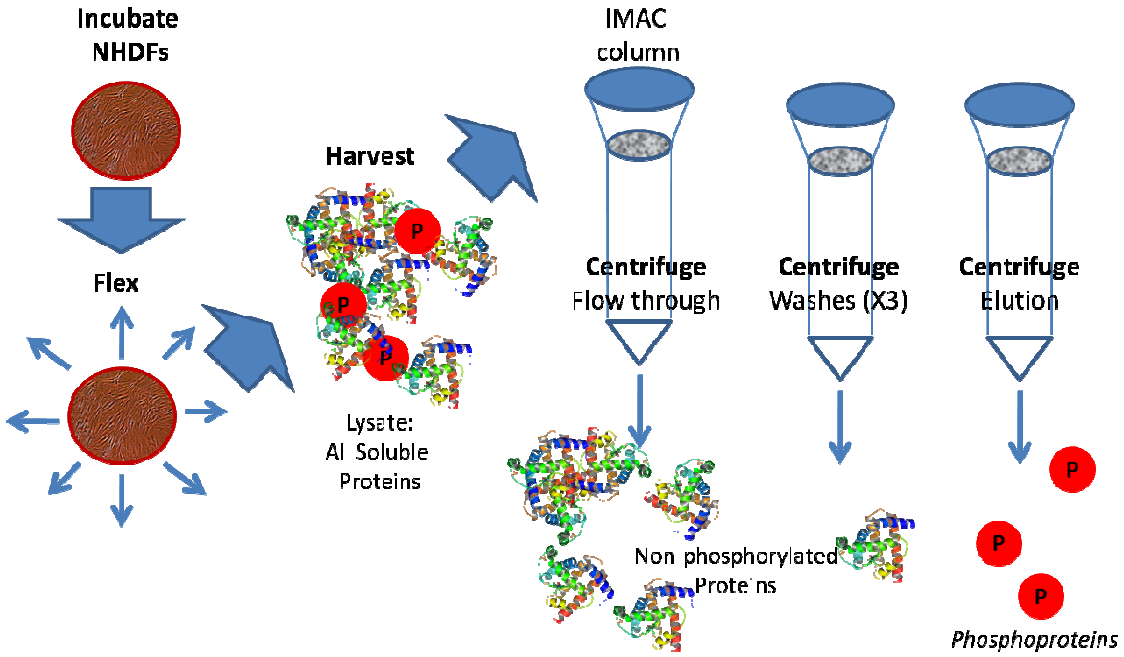

Figure 3.1 Phosphoprotein enrichment

These cellular samples contain information on the short term response to external stimuli within the intracellular signaling proteins. A relevant subset of signaling proteins are phosphorylated proteins, or phosphoproteins. These proteins are "activated"; their function is altered by the addition of one or more phosphate groups. Oftentimes the function of a phosphorylated protein is to cause the phosphorylation or dephosphorylation of another protein. Enrichment for phosphoproteins is possible and necessary for mass spectrometry, as discussed later. A specified amount of total protein is loaded onto a phosphoprotein enrichment column. The current preferred technique is immobilized metal affinity chromatography (IMAC). This enriches for all types of phosphoproteins by binding with the phosphate groups. It is important to note that enrichment techniques are not completely specific for their target proteins; other proteins may be included through non-specific binding or some target proteins may not be included. Free phosphates, which are not bound to proteins, could interfere with this enrichment method,

should be removed by the rinse step during harvest. This is a source of potential variation or noise. If non-phosphorylated proteins bind, it is not know whether they would contribute to the analysis (i.e. provide useful information) or interfere with it. Free phosphates may prevent phosphoproteins from binding to the column; this is why it is crucial that the cells are thoroughly rinsed at harvest.

Mass spectrometry of Intact Phosphoproteins

The subsequent enrichments can be evaluated by mass spectrometry. Matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry is employed to measure the mass profile of the enriched sample.

(All reagents were purchased from Sigma-Aldrich unless otherwise noted.) The spotting/dilution solution is 50:50 acetonitrile (ACN) and 0.1 % trifluoroacetic acid (TFA) in water and is made fresh weekly. A sinapinic acid matrix, comprising saturated sinapinic acid in spotting/dilution solution, is used within a day (a glass pipette is used to obtain a dry aliquot of sinapinic acid and is added to 0.5 mL of spotting/dilution solution, vortex mixture well, centrifuge, and use supernatant fluid). Samples are desalted using a ZipTip® (Millipore); a volume of 1.0 µL (3.3 µg total protein) of eluate is mixed with 1.0 µL of the matrix and dried on a metal target plate. Standards that have m/z values close to the range of interest are also spotted; cytochrome C and bovine serum albumin (Thermo Scientific) were used.

Target plate was loaded in a Reflex III MALDI-TOF (Bruker). Laser settings were optimized for each sample type. Replicate spectra were obtained from each sample or standard spot. Bruker data processing tools were used to correct baseline and reduce noise.

Phosphopeptide Analysis

Cell Lysis

Cells that were to be analyzed for phosphopeptides were harvested in an adaptation of the Gygi protocol (Villén and Gygi, 2008). Lysis buffer of 8 M urea, 75 mM NaCl, 50 mM Tris pH 8.2, 1 mM phenylmethanesulfonyl fluoride (PMSF), and Halt® Protease and Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific) was prepared a few hours before harvest; the inhibitor cocktail and PMSF were added immediately prior to harvest. Culture medium was removed, cells were rinsed with ice-cold 50 mM HEPES and a second rinse with ice-cold serum-free

medium. Cells were lysed in the plate/flask with 1 mL per 175 cm^2 or proportional volume to growth area.

Peptide phosphotyrosine immunoprecipitation

The tryptic peptides are incubated with a pan-specific antibody for phosphotyrosine, and then eluted at low pH. The eluted phosphotyrosine enrichment is either cleaned of salts with a STAGE tip or passed through IMAC for a second round of enrichment to lessen the carryover of non-specifically bound peptides. Then phosphotyrosine -enriched peptides are analyzed by LC-MSMS.

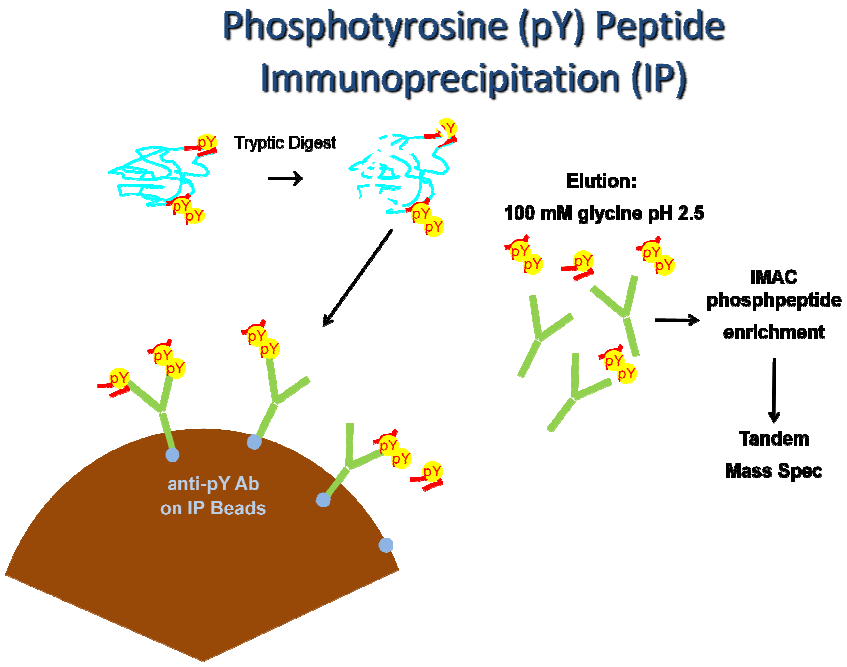

Figure 3.2: Technique for targeting phosphotyrosine.

Tandem Mass Spectrometry

Identification

The software MaxQuant (Cox and Mann, 2008) was designed for isotope-labeled samples. It detects peaks, isotope clusters and stable amino acid isotope labeled (SILAC) peptide pairs as three-dimensional objects in m/z, elution time and signal intensity space. By integrating

multiple mass measurements, mass accuracy in the parts per billion range is achieved. MaxQuant quantifies several hundred thousand peptides per SILAC proteome experiment. Newer versions of MaxQuant have been included label-free quantification and Andromeda (Cox et al., 2011), an internal database search for identification.

Quantification

Raw profile files were converted to centroid mzxml format with msConvert (Kessner et al., 2008) peak picking TRUE. These mzxml files were simultaneously analyzed by Sue Van Riper's algorithm, proximity-based intensity normalization (PIN). PIN is part of a larger software framework termed RIPPER (Van Riper et al. submitted). PIN is designed to normalize intensities for detection of biological variation in LC-MSMS data. This algorithm extracts valid peptide signals, sums intensities in the isotopic envelope (similar accuracy as integrating area under the curve, but with faster processing), and performs local normalization of intensities within and across samples.

Response

mRNA Quantification

RNA was analyzed quantitatively for collagen and elastin. RNA purification was performed using the RNeasy Mini or RNeasy Plus Mini (Qiagen) kits according to manufacturer's protocols. RNA was quantified via NanoDrop (Thermo Scientific); 60 ng of RNA was reversed transcribed to cDNA using SuperScript Reverse Transcriptase III (Invitrogen). A control without reverse transcriptase was performed for each sample.

Quantitative Real Time Polymerase Chain Reaction

Primers for RPL22 and/or GAPDH were used as reference genes for each real time analysis. Primers for COL1A1 and Elastin were used. A control without template was included for each primer set. Using a SYBR green (Stratagene) master mix with reference dye ROX at 1:50 dilution, samples were placed in triplicate on a 384-well plate and analyzed with an Applied Biosystems 4900 HT. The ΔΔCt method (Livak and Schmittgen, 2001) was used to obtain fold change in mRNA expression.

Extracellular Matrix Assays

Absolute amounts of collagen and elastin in the tissue can be measured. Conditioned medium was sampled from the same tissue culture dish throughout the course of each experiment; data from these are paired. A hydroxyproline assay and an ELISA for procollagen type I C-peptide (PIP) were used to determine deposited and secreted collagen, respectively. A ninhydrin assay was used to determine deposited elastin. A Hoechst assay for DNA content was employed to determine cellularity.

Lysyl Oxidase Activity

Lysyl oxidase (LOX) activity can be monitored fluorometrically (Palamakumbura and Trackman, 2002). Hydrogen peroxide (H_2O_2) is produced during the enzymatic reaction of LOX with its substrate. An *in vitro* substrate is cadaverine dihydrochloride. Amplex® UltraRed reagent reacts with H_2O_2 and horseradish peroxidase to yield a fluorescent product. Conditioned medium was assayed directly for lysyl oxidase activity and indirectly for collagen secretion. Standards were made with or without medium that was identical to the conditioned medium samples except that it had never contacted cells. Amplex® UltraRed (Invitrogen) measures H_2O_2 production; LOX-produced H_2O_2 is distinguished by analyzing samples with and without the LOX inhibitor β-aminopropionitrile (BAPN). The bench top protocol is a combination of the Amplex® UltraRed manual and the method developed by Palamakumbura and Trackman; it is still being modified within the Tranquillo Laboratory. Briefly, samples of conditioned medium and medium that had not contacted cells were thawed on ice. A stock buffer of 3.6 M urea, 150 mM sodium borate was made fresh by heating and stirring, then cooled to room temperature and brought to pH 8.2 with HCl. H_2O_2 was added in known concentrations to create a standard curve. The diluent for the standard curve is medium of the same composition and concentration as the conditioned medium samples, only that it has not contacted cells.

Data Analysis

Data from samples taken immediately prior to perturbation can be used as a zero time point in temporal profiles, or for a common reference for fold changes of corresponding post perturbation phosphopeptide signals, mRNA expression, soluble collagen, and LOX activity. The number of distinct phosphopeptides common across all samples is typically in the hundreds; a

situation of high number of variables and small sample number. LC-MSMS data is inherently discovery-based. This suits the project well as there is no complete mechanistic model of intracellular signaling proteins involved in tissue growth. The response data, however, are all known targets that are well described. Soluble and insoluble collagen amounts are quantified by absolute amount (e.g. ng collagen). Currently, all other sample types are analyzed using relative quantification. It is crucial that the quantification be at a level that comparisons can be made across separate experiments performed at different times. A relative metric of more collagen production in one treatment versus another may be sufficient for defining R. Upon further investigation, it may be found that absolute metrics will be necessary for some or all of these data types.

Chapter 4 Systems Biology Results and Discussion

The focus of this project was to obtain information from intracellular signaling proteins predictive of tissue growth. The first step is to establish a protocol that can elucidate differences between phosphoproteome profiles generated by different treatments.

Example Experiment

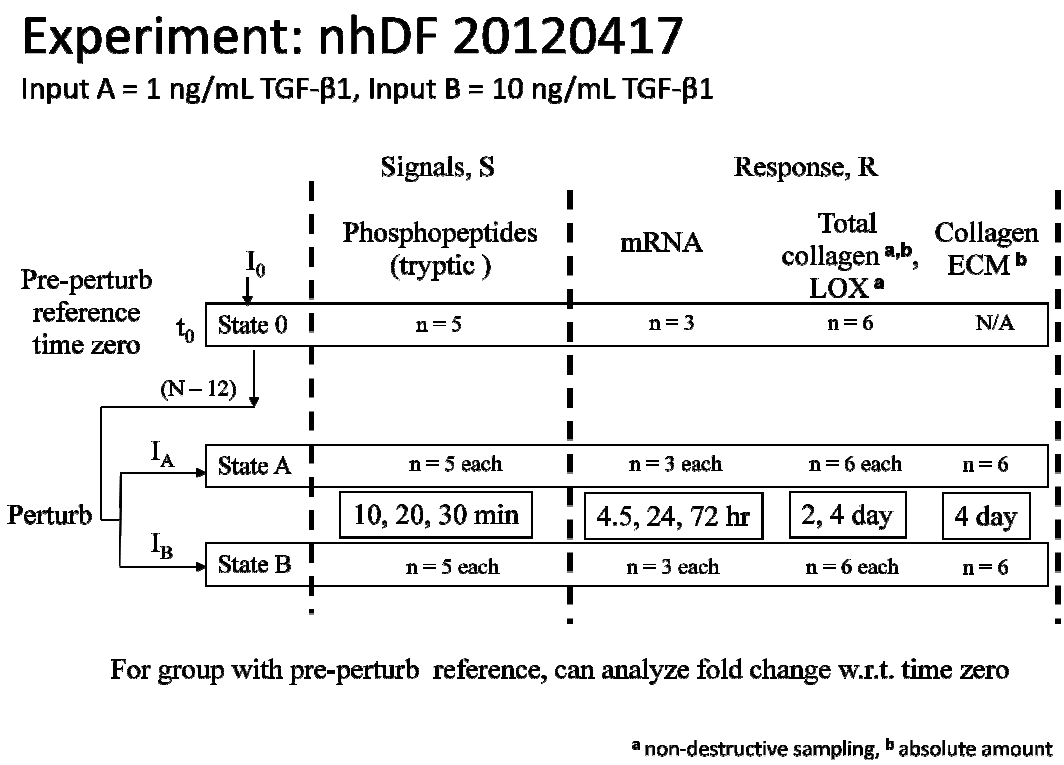

Figure 4.1: Experimental workflow

Multiple Time Points

It is currently unknown what time points contain relevant information predictive of downstream tissue growth. Phosphorylation changes have been observed within the first 30 minutes of perturbation.

Pre-Perturbation Reference

The change from a pre-perturbed suboptimal state to a post-perturbed state could contain useful information. This could serve as a reference for experiment-to-experiment comparisons. Pre-perturbed suboptimal states may vary at the systems level for each perturbation.

Converting from one suboptimal state versus a distinct suboptimal state may require different participation of phosphosites, for example.

Methods

Neonatal human dermal fibroblasts (nhDF) were expanded in 6-well plates from 8,500 cells/cm^2 to near confluence, at which point basal medium (DMEM w/o phenol red, 10% FBS, 100 U/ml penicillin, 100 µg/ml streptomycin, 2 µg/mL insulin, 50 µg/mL ascorbic acid) for tissue production was applied. (It was later found that this medium was lacking L-glutamine, which is generally helpful for tissue growth.) After two days, conditioned medium was collected and a portion of the cells were harvested for intracellular proteins or mRNA as pre-perturbation references.

The remaining nhDF were perturbed with fresh basal medium supplemented with 1 or 10 ng/ml TGF-β1. For harvesting at multiple time points, the beginning of each perturbation was staggered to allow sufficient time for harvest (Table 4.1). Cells were lysed at 10, 20, or 30 minutes for intracellular proteins. Cells were harvested for mRNA at 4.5, 24, and 72 hours post perturbation. 48 and 96 hrs after perturbation, conditioned medium was sampled from the same cell plates as the pre-perturbation references (fresh perturbation medium was applied after the 48 hr sampling).

Plates	@ 37 deg C	Time point end	Rinses + Lysis Buffer	@ 4 deg C
Z0 (2 plates)	6:47	6:47	6:57, 6:57	6:58
A1 & B1 (4 plates)	7:10	7:20	7:28, 7:34	7:35
A2 & B2 (4 plates)	7:15	7:30	7:41, 7:47	7:48
A3 & B3 (4 plates)	7:20	7:50	7:58, 8:04	8:05

Table 4.1 Times (hr:min PM) of several key steps of the Signals perturbation and harvest.

Intracellular proteins lysates were sonicated on ice, and their protein concentrations were determined using the BCA assay. 100 µg from each sample was digested to peptides via trypsin and washed of salts. Iron Immobilized Metal Affinity Chromatography (IMAC) was employed to enrich for phosphopeptides. Phosphopeptide enrichments were analyzed by tandem mass spectrometry on an Orbitrap Velos in HCD profile mode in the Center for Mass Spectrometry

and Proteomics (University of Minnesota). There were 35 samples (replicates of 5 for pre-perturbation reference, 1 or 10 ng/ml TGF-β1 at 10, 20, or 30 min); no fractionation prior to LC-MSMS was performed, thus there were 35 runs on the mass spectrometer.

Results

Signals

Identification

LC-MSMS raw files (35 files) were loaded into MaxQuant version 1.3.0.5. Parameters: carbamidomethyl of cysteine as a fixed modification, variable modifications of phosphorylation (on serine, threonine, or tyrosine), oxidation of methionine, and N-terminus protein acetylation; UniProt human database. MaxQuant identified 2894 phosphosites (2341 serines, 511 threonines, and 42 tyrosines).

Quantification

LC-MSMS raw files were converted to mzxml and analyzed using PIN developed here at the University of Minnesota (Van Riper et al., submitted), resulting in 7205 peptide signals quantified. Peptide signals that matched phosphopeptides identified by MaxQuant (within +/- 0.002 m/z and having the same charge) were selected for further analysis. Different charge states for the same peptide had their normalized intensities summed. A total of 1689 phosphopeptides were both identified by Andromeda (Cox et al., 2011) and quantified using PIN.

Statistical Analysis

Phosphopeptide signals were further analyzed if they had 3 or more nonzero intensities (out of 5 intensities) in at least one time-treatment group of 1 or 10 ng/mL TGF-β1 (at 10, 20, or 30 min), resulting in 903 phosphopeptide signals. Two-part linear-logistic models (Duan et al., 1983) were employed. For each phosphopeptide signal, a logistic regression is performed on all the 30 observations (5 replicates of three time points each for 1 or 10 ng/mL TGF-β1). Then χ^2-tests were performed to determine the significance of time (as a factor), treatment and the time-treatment interaction. The second part of the two-part model was to fit a linear model for each phosphopeptide signal, using only the nonzero intensities. Overall F-tests were performed,

comparing the fitted model to the null model, which had no time or treatment effect or interaction.

Treating time as a factor, we found no significant peptide signals using the Benjamini-Hochberg procedure with level 0.05. However, at level 0.1, we found five significant overall F-tests for the linear part. That means the first five smallest p-values were smaller than $i \times 0.1/903$, $i = 1, 2, \ldots, 5$, respectively. Their masses are 948.443, 1223.5799, 1573.6708, and 1574.7454 (two assigned quantitative profiles) Da.

For the logistic regression part, at level 0.1, we found nine significant χ^2-tests for treatment effect but no significant χ^2-test for time effect or interactions. Their masses are 864.37827, 1097.5118, 1233.4486, 1310.6305, 1389.4948, 1438.7433, 1453.6844, 1574.7454 (quantitative profile is unique from the two found by the linear part), and 1677.824 Da.

We then repeated this two-part model fitting by treating time as numeric values (1, 2, 3) indicating 10, 20, and 30 minutes. For the linear part, we performed F-test of overall effects and the p-values with the Benjamini-Hochberg procedure give 8 significant peptide signals at a level of 0.05. They have masses of 948.443, 1201.5856, 1223.5799, 1325.6414, 1360.5483, 1574.7454 (quantitative profiles are identical to the two found by the time as factor linear model), 2580.0006 Da. One additional signal was found at a level of 0.1, having a mass of 2580.0006 Da (quantitative profile is unique from the one found at 0.05). Four of them were also significant peptide signals found by treating time as a factor.

For the logistic regression part, at level 0.05, we found one significant χ^2-test for time effect (matching to reverse sequence decoy HTFSGVAS(ph)VES(ph)SS(ph)GEAFHVGK, 2358.896 Da) and at level 0.1, we have two more significant peptide signals (1394.6999 and 3522.4501 Da). At level 0.05, we found four significant χ^2-tests for treatment effect (1097.5118, 1233.4486, 1438.7433, and 1453.6844 Da); at level 0.1, we have three more (masses of 1310.6305, 1389.4948, and 1677.824 Da). No significant χ^2-test for interaction effect was found.

In total we found 22 peptide signals significant for time/treatment effect, shown in Table 4.2 by increasing mass in Daltons (Da). The sequences GGVLS(ph)PSPR and HTFSGVAS(ph)VES(ph)SS(ph)GEAFHVGK are decoy reverse sequences, suggesting a 2/22=9.1% false discovery rate with this analysis. For some masses, there were multiple quantitative

profiles (from PIN) that matched to the same m/z (within ± 0.002) (rows with blank cells in Table 4.2).

There were also multiple identifications that matched within ± 0.002 m/z for some masses (Table 4.3). Each of these can have their corresponding fragment spectra manually evaluated for verifying sequence and/or phosphosite location, which has not been done yet. For example, the VQIPVSRPDPEPVSDNEEDSYDEEIHDPR sequence has evidence for the location of the two phosphorylations; if the peptide was fragmented within the mass spectrometer between the indicated potential serine (S) and tyrosine (Y) phosphorylations, then the existence of one or more phosphoforms can be determined. Likewise the 864.37827 Da peptide signal matches the sequences LSGFS(ph)FK or SFGLFS(ph)K; this ambiguity in the location of the leucine (L) may be discerned if sufficient fragmentation occurred within the SGF sequence.

According to UniProt and Ingenuity Pathway Analysis, many of the identified proteins are located in the cytoplasm and/or cytoskeleton, with functions including kinase activity, signaling, regulation, and organization (Table 4.4).

MASS (Da)	MODIFIED SEQUENCE	PROTEIN NAME
864.37827	LSGFS(ph)FK*	Myristoylated alanine-rich C-kinase substrate*
948.443	GGVLS(ph)PSPR	[REVERSE SEQUENCE]
1097.5118	GSLAS(ph)LDSLR*	Catenin delta-1*
1201.5856	LPS(ph)APSGGAPIR	Obscurin
1223.5799	(ac)AS(ph)GVAVSDGVIK	Cofilin-1
1233.4486	RDYDDMS(ph)PR	Heterogeneous nuclear ribonucleoprotein K
1310.6305	M(ox)LISAVS(ph)PEIR	Transcription factor E2F8
1325.6414	KQIT(ph)MEELVR	Plectin
1360.5483	GEPNVSY(ph)ICSR	Glycogen synthase kinase-3 alpha/beta
1389.4948	YFDS(ph)GDYNMAK	Endosulfine alpha
1394.6999	S(ph)LPVSVPVWGFK	Proline-rich AKT1 substrate 1
1438.7433	KIS(ph)GTTALQEALK	CAP-Gly domain-containing linker protein ½
1453.6844	T(ph)PRTPRT(ph)PQLK	La-related protein 1
1573.6708	MFGGPGT(ph)ASRPSSSR	Vimentin
1574.7454	TYS(ph)LGSALRPSTSR*	Vimentin*
1574.7454		
1574.7454		
1677.824	RLS(ph)EQLAHTPTAFK	Phosphatidylinositol 4-kinase beta
2358.896	HTFSGVAS(ph)VES(ph)SS(ph)GEAFHVGK	[REVERSE SEQUENCE]
2580.0006	MGPSGGEGM(ox)EPERRDS(ph)QDGSSYR*	LIM and SH3 domain protein 1
2580.0006		
3522.4501	VQIPVSRPDPEPVS(ph)DNEEDSY(ph)DEEIHDPR*	Tight junction protein ZO-1

Table 4.2: Peptide Signals ordered by mass that were significant between 1 and 10 ng/mL TGF-β1 or across time, with matching identifications. Rows with blank cells indicate multiple quantitative profiles matched to the same mass. *Alternative matches exist for sequences/proteins within the ± 0.002 m/z tolerance for matching.

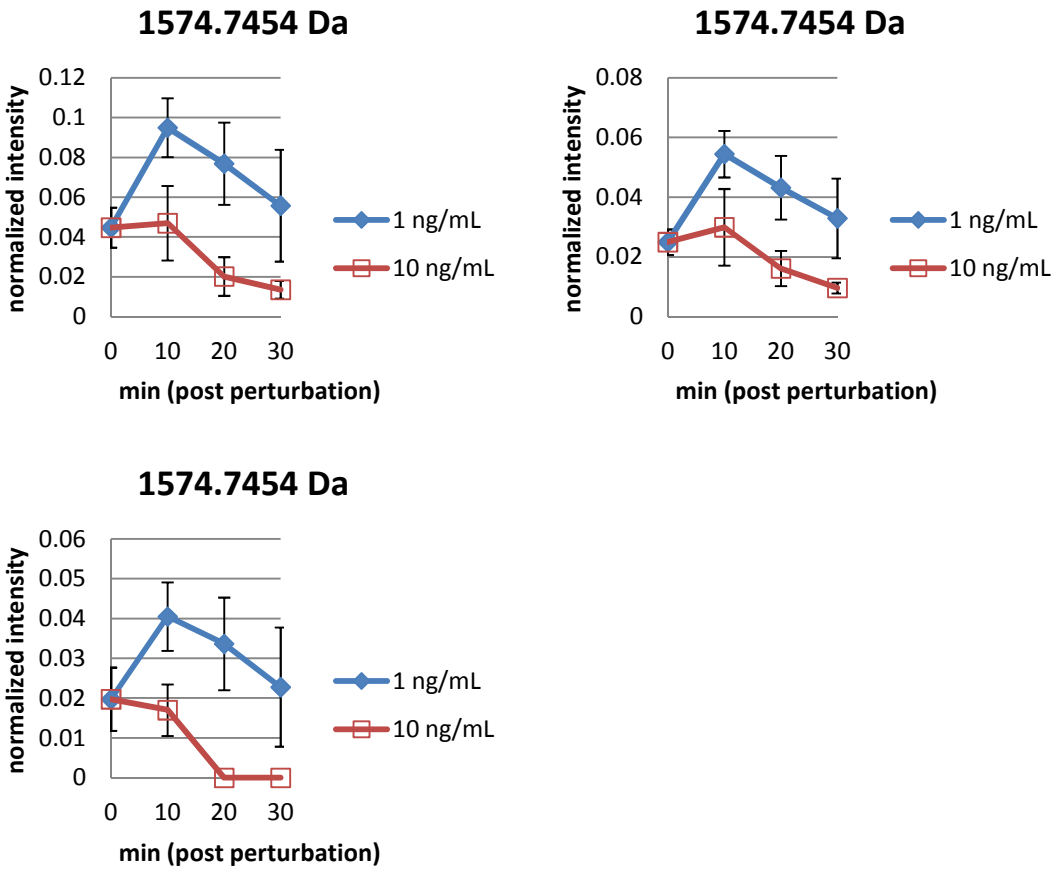

Figure 4.2: Temporal profiles of three significant peptide signal profiles that match 1574.7454 Da within the ± 0.002 m/z tolerance for matching.

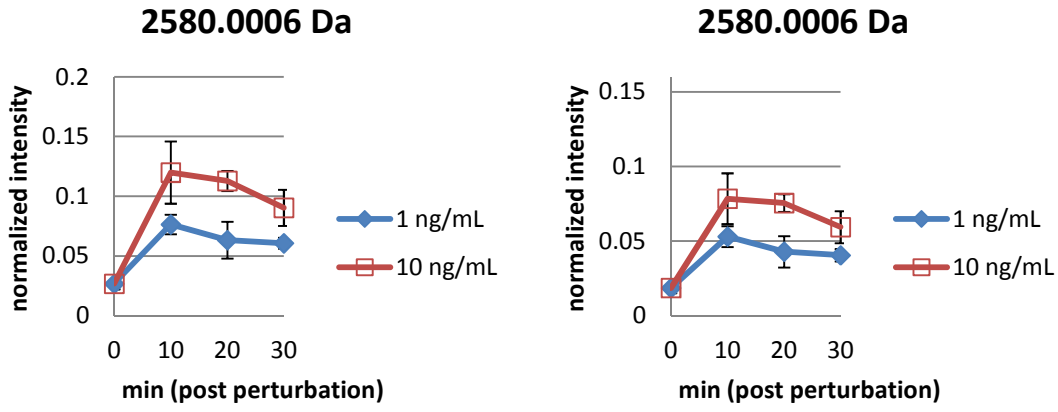

Figure 4.3: Temporal profiles of two significant peptide signal profiles that match 2580.0006 Da within the ± 0.002 m/z tolerance for matching.

MASS (Da)	MODIFIED SEQUENCE	GENE
864.37827	LSGFS(ph)FK	MARCKS
864.37827	SFGLFS(ph)K	EPG5
1097.5118	GSLAS(ph)LDSLR	CTNND1
1097.5118	GSS(ph)LKIEER	CTNND1
1574.7454	TYS(ph)LGSALRPSTSR	VIM
1574.7454	TYSLGSALRPS(ph)TSR	VIM
1574.7454	KEEQNKS(ph)HPISAK	DST
3522.4501	VQIPVSRPDPEPVS(ph)DNEEDSY(ph)DEEIHDPR	TJP1
3522.4501	VQIPVS(ph)RPDPEPVS(ph)DNEEDSYDEEIHDPR	TJP1
3522.4501	VQIPVSRPDPEPVS(ph)DNEEDS(ph)YDEEIHDPR	TJP1
3522.4501	VQIPVSRPDPEPVSDNEEDS(ph)Y(ph)DEEIHDPR	TJP1

Table 4.3: Alternative matches for sequences/proteins within the ± 0.002 m/z tolerance for matching. (Shading is solely for the purpose of separating each mass.)

Gene	PROTEIN NAMES	Function	Location
EPG5	Ectopic P granules protein 5 homolog	autophagy	Unknown
MARCKS	Myristoylated alanine-rich C-kinase substrate	Substrate for protein kinase C. This protein binds calmodulin, actin, and synapsin; filamentous (F) actin cross-linking protein.	Cytoplasm › cytoskeleton. Membrane; Lipid-anchor
CTNND1	Catenin delta-1	Wnt signaling pathway. May associate with and regulate the cell adhesion properties of both C- and E-cadherins.	Cytoplasm. Nucleus. Cell membrane.
CALD1	Caldesmon	movement, cross-linkage, translocation, morphology, cell cycle progression, assembly, induction, formation	actin cytoskeleton, plasma membrane
OBSCN	Obscurin	organization, structure, apoptosis; regulation of Rho protein signal transduction, kinase activity	intracellular
CFL1	Cofilin-1	organization, size, depolymerization, migration, apoptosis, stabilization, formation, motility, cytokinesis	actin cytoskeleton
HNRNPK	Heterogeneous nuclear ribonucleoprotein K	role in p53/TP53 response to DNA damage,	Cytoplasm. Nucleus › nucleoplasm.
E2F8	Transcription factor E2F8	transcription repressor, preventing p53/TP53-dependent apoptosis	Nucleus
PLEC	Plectin	Anchors intermediate filaments to desmosomes or hemidesmosomes.	Cytoplasm › cytoskeleton. Cell junction › hemidesmosome
GSK3B	Glycogen synthase kinase-3 alpha/beta	Constitutively active protein kinase, glucose homeostasis,	Cytoplasm. Nucleus. Cell membrane.
ENSA	Endosulfine alpha	response, binding	Cytoplasm
AKT1S1	Proline-rich AKT1 substrate 1	differentiation, size, apoptosis, growth, autophagy, biogenesis; negative regulation of TOR signaling cascade	Cytoplasm, cytosol, Nucleus
CLIP1/2	CAP-Gly domain-containing linker protein 1/2	Seems to link microtubules to dendritic lamellar body (DLB)	Cytoplasm › cytoskeleton
LARP1	La-related protein 1	positive regulation of macroautophagy	Cytoplasm
VIM	Vimentin	stabilization of type I collagen mRNAs for CO1A1 and CO1A2, class-III intermediate filaments found in various non-epithelial cells	Cytoplasm
DST	Dystonin	organization, motility, endoplasmic reticulum stress response, cell viability, cell death, coalignment; actin/ATP binding	cytoplasm, integral to membrane, microtubule cytoskeleton, nucleus
PI4KB	Phosphatidylinositol 4-kinase beta	Phosphorylates phosphatidylinositol in the second messenger inositol-1,4,5,-trisphosphate (PIP).	mitochondrion, RER. Golgi. Cytoplasm › perinuclear region.
LASP1	LIM and SH3 domain protein 1	regulation of dynamic actin-based, cytoskeletal activities.	Cytoplasm › cell cortex, cytoskeleton
TJP1	Tight junction protein ZO-1	assembly, barrier function, development, quantity, stabilization, density, adhesion, proliferation, invasion	cytoplasm, plasma membrane, tight junction

Table 4.4: Biological functions for assigned identifications of the significant phosphopeptides (information from UniProt and Ingenuity Pathway Analysis).

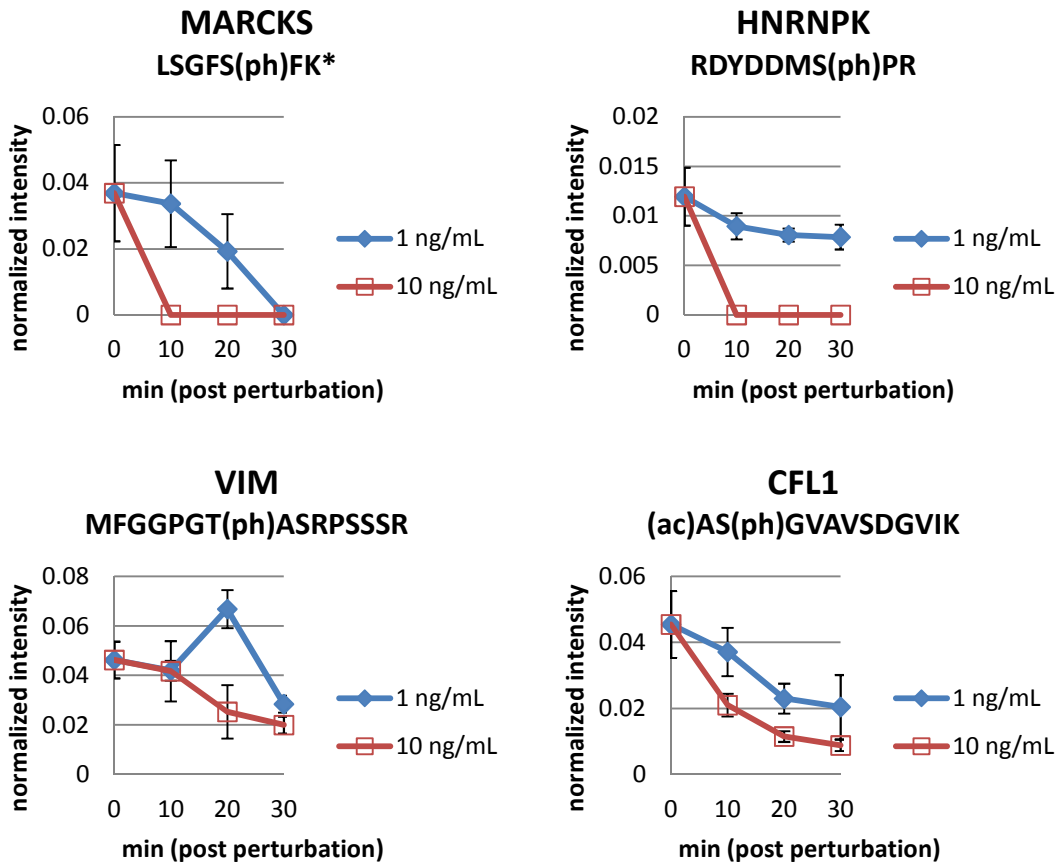

Figure 4.4: Temporal profiles of four significant phosphopeptides that trend down with time. *The mass for LSGFS(ph)FK also matched to SFGLFS(ph)K belonging to EPG5.

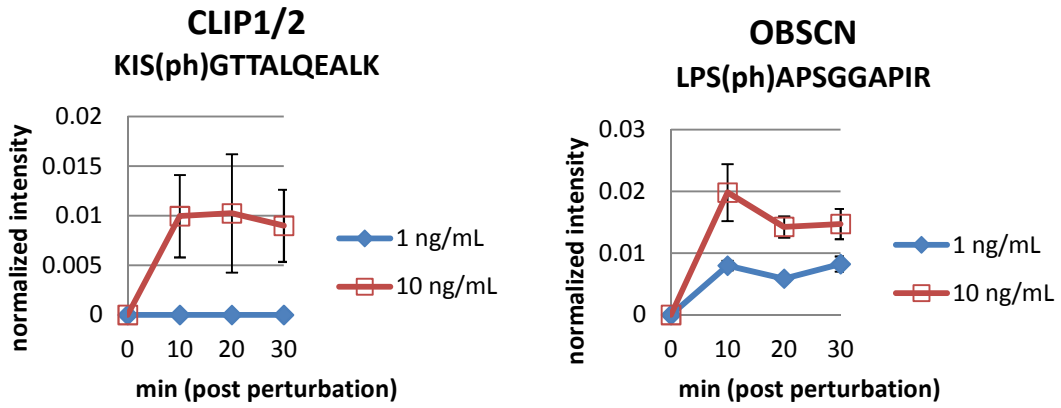

Figure 4.5: Temporal profiles of two significant phosphopeptides that trend up with time.

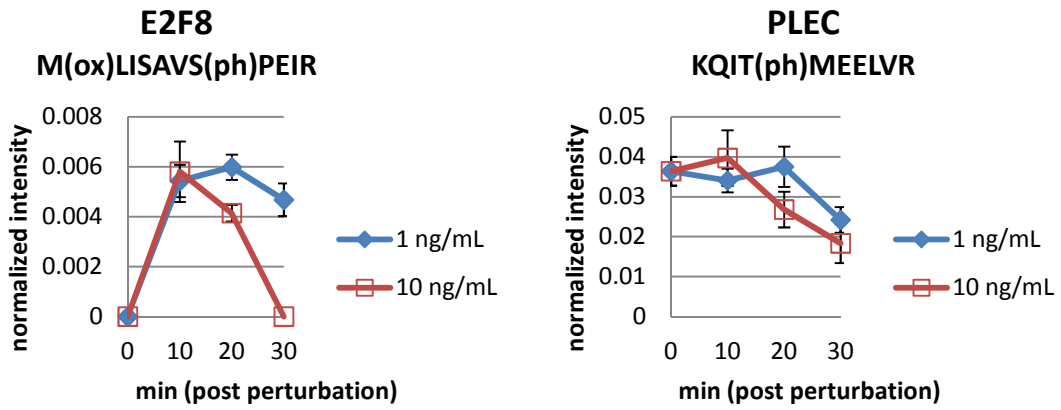

Figure 4.6: Temporal profiles of significant phosphopeptides showing various patterns.

Conclusions

In general, those peptide signals which decreased from 0 to 10 minutes had the 10 ng/mL TGF-β1 profile on bottom (Figure 4.4) and those peptide signals which increased from 0 to 10 minutes had the 10 ng/mL TGF-β1 profile on top (Figure 4.5).

Phosphopeptide Enrichment Efficiency

To quantitatively determine phosphopeptide enrichment efficiency, the sum of peptide signal intensities of all phosphopeptides (identified by MaxQuant) for a sample was divided by the sum of peptide signal intensities of all peptides for the same sample.

$$\frac{\sum_{i=0}^{p} I_i^{ph}}{\sum_{j=0}^{m} I_j} * 100\%$$

Equation 4.1 Where I_i^{ph} is the intensity of the i^{th} phosphopeptide (identified by MaxQuant and quantified by PIN), I_j is the intensity of the j^{th} peptide signal (quantified by PIN; identification by MaxQuant only for phosphopeptides), p is the number of phosphopeptides and m is the number of all peptides ($m \geq p$) in the sample/run.

Replicates 1 and 2 were trypsin digested, enriched for phosphopeptides, and analyzed via LC-MSMS on an Orbitrap Velos. Replicates 3, 4, and 5 were treated in the same manner, but 4 months later. The samples of replicates 1 and 2 had enrichment efficiency (% phosphopeptide intensities of all peptide intensities, based on PROXINORM) of 60 ± 7 % (mean ± standard deviation), while the samples of replicates 3, 4, and 5 had an enrichment efficiency of 36 ± 4 %

(mean ± standard deviation). A two-tailed Student's T-test assuming equal variance between these two data subsets resulted in a p-value less than $5x10^{-15}$, indicating that the early sample preparation and analysis was quantitatively more efficient for phosphopeptides (Figure 4.7). The sum of all peptide signal intensities (denominator in Equation 4.1) did not differ significantly between these two subsets. The sum of phosphopeptide intensities (numerator in Equation 4.1) differed ($p<5x10^{-9}$).

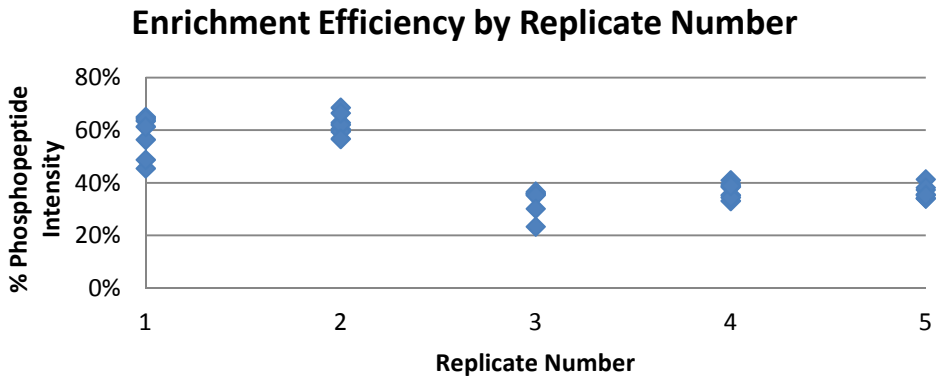

Figure 4.7 Phosphopeptide enrichment efficiency determined quantitatively via Equation 4.1 across replicate number.

Response

Samples for mRNA (n>3 for each group) were purified for RNA; 60 ng aliquots were reverse-transcribed to cDNA. The cDNA was analyzed via qPCR with COL1A1 and ELN primers (RPL22 as a reference gene). Fold changes in expression were significant across time (4.5, 24, and 72 hr), but not between treatments (1 or 10 ng/ml TGF-β1) (Figure 4.8).

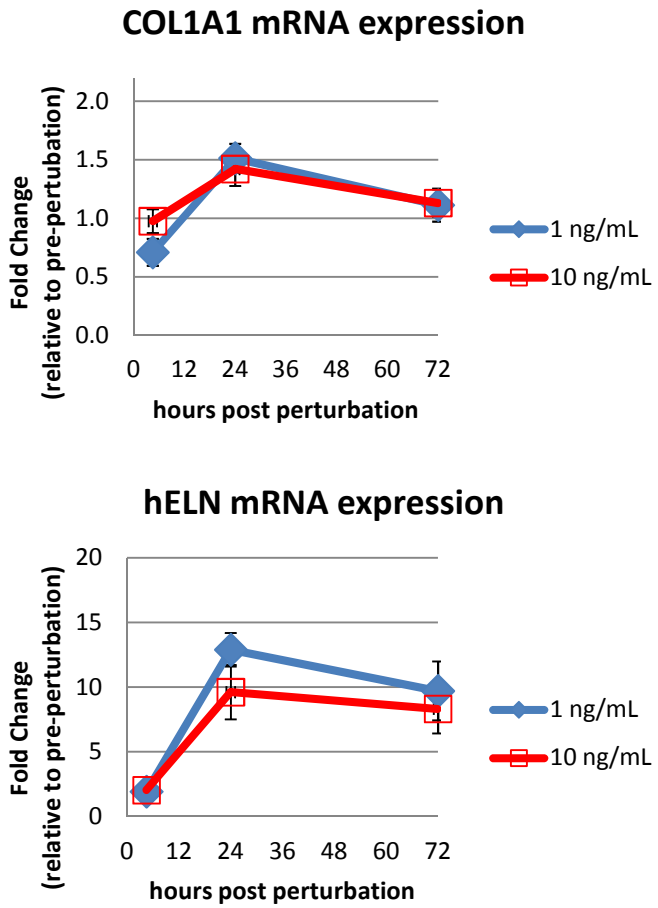

Figure 4.8: Temporal profiles of collagen type 1 (COL1A1, top) and elastin (hELN, bottom)

In the conditioned medium samples (n>3 for each group), LOX activity was determined by a fluorometric assay and an indicator of collagen I secretion was assayed in a PIP ELISA (Figure 4.9); no significant differences were found between groups in either medium assay.

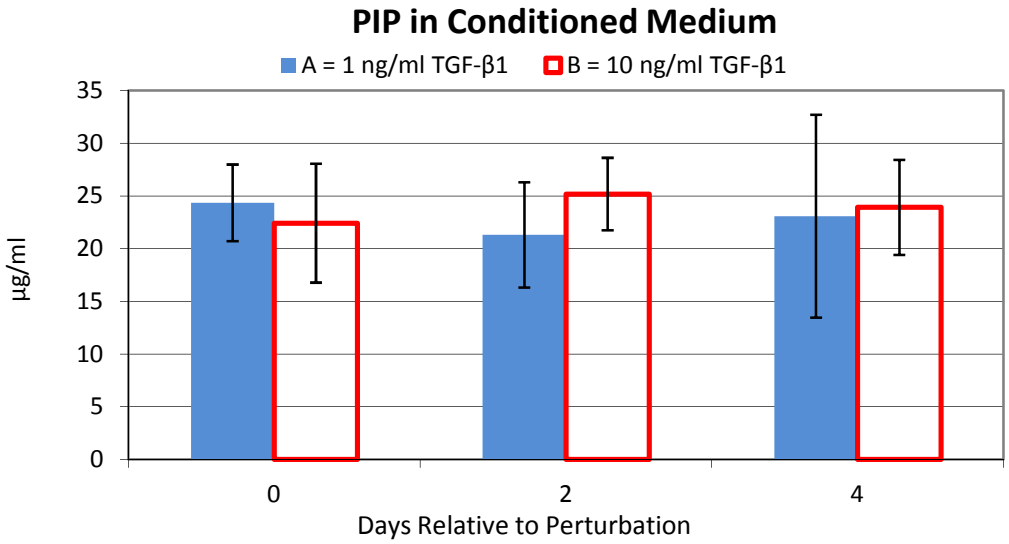

Figure 4.9: PIP ELISA results for conditioned medium samples. Sandra Johnson ran the ELISA assay and data analysis.

The 4 day cell layer samples for DNA and ECM assays had six samples per treatment group. Cell and elastin content per cell did not differ between treatments (Figure 4.10 upper left and right panels, respectively). Collagen increased ($p<0.005$) from 13.0 ± 1.3 µg per million cells to 16.3 ± 1.5 µg per million cells (Figure 4.10 lower left panel). Total protein also increased ($p<0.005$) (Figure 4.10 lower right panel).

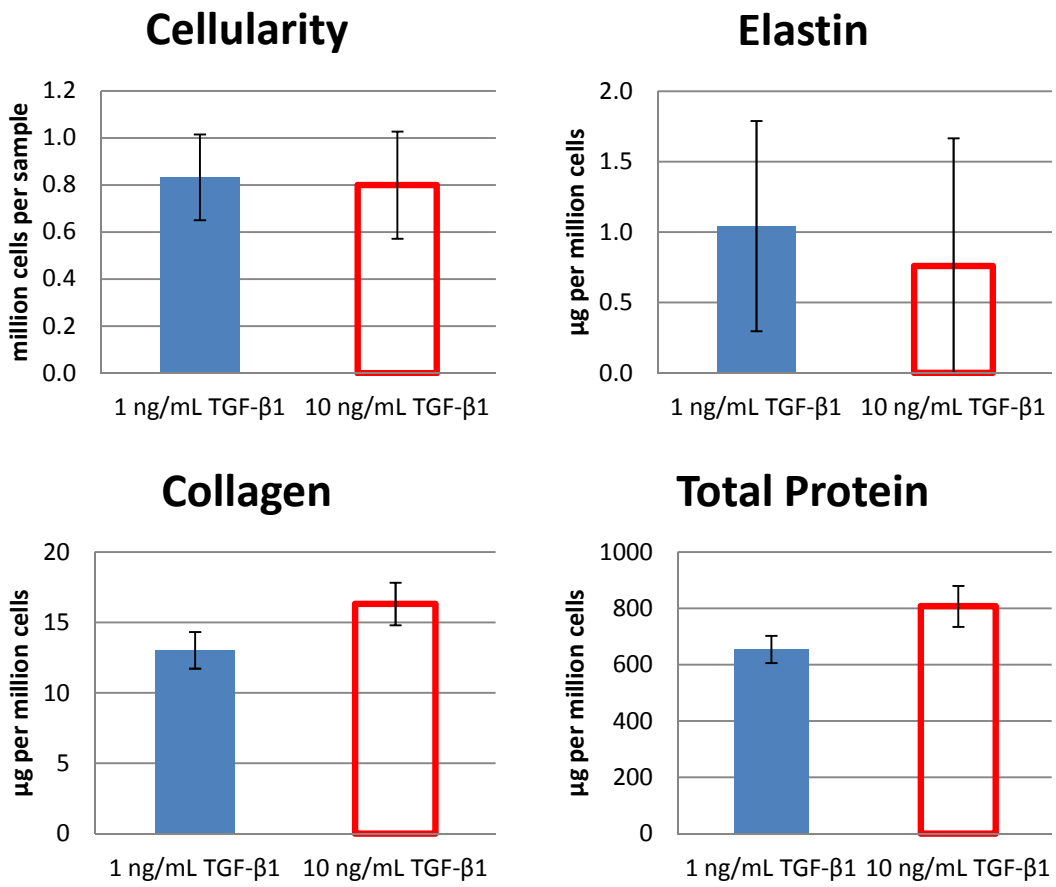

Figure 4.10: Tissue data 4 days post perturbation. Top left panel: Cells per sample in millions; top right panel elastin content per million cells; bottom left panel: collagen content per million cells; bottom right panel: total protein content per million cells.

Discussion

If the trends seen in the signals data of the Example Experiment are consistent in future work, biologically relevant interpretation can be made. Because peptide signals which decreased from 0 to 10 minutes had the 10 ng/mL TGF-β1 profile on bottom (Figure 4.4), they indicate that certain phosphoproteins are inhibited by increasing TGF-β1. Other peptide signals, which increased from 0 to 10 minutes and had the 10 ng/mL TGF-β1 profile on top (Figure 4.5), were increased with a higher concentration of TGF-β1. If the association of these signals trends to downstream collagen response is verified, the activation and inhibition of specific groups of

phosphopeptide signals achieves a greater collagen deposition due to the higher concentration of TGF-β1 (Figure 4.10).

Overall, the quantitative data needed for Signals (S) & Response (R) can be acquired. Validation, including further experimentation is needed for S data. The methods developed can be used in finding R~f(S).

Recommendations for Further Work

The significant peptide signals must be verified both for identification and quantification. This can be done by further experimentation and utilizing more of the data available. The data contain, for example, connection between the MS precursor mass (matched to quantitative profile(s)) and the associated fragment spectra (necessary for identification). This can be accessed for verifying matches of quantification and identification. LC retention time can also be used to aid the matching. The current analysis already sums different charge states of the same peptide, as these are biologically the same analyte, even though the biochemical analysis treats them as different. For the same reason, missed cleavages and/or overlapping sequences are biologically the same; therefore further improvement could include a computational method for summing these intensities.

A single phosphorylation in a peptide can be regulated differently than other phosphorylation(s) in the same peptide, so different phospho-forms of the same amino acid sequence may have levels of a similar function or divergent functions. The current analysis does not distinguish between phosphor-forms, but this information may be able to be obtained from the existing data.

Once a model R~f(S) is described, targeted methods can be used to quantify the important signals. Targeted mass spectrometry, such as selective reaction monitoring (SRM) (Gallien et al., 2011), may be particularly useful. Techniques that do not use mass spectrometry could also be employed.

The final protein composition in the ECM is of primary interest. Mass spectrometry is advantageous for investigating post-translational modifications of these proteins. A SRM method would have good potential to target collagen, elastin, lysyl oxidase, and other secreted proteins simultaneously in a sample. Synthetic versions of target peptides containing stable-

isotopes can be used for absolute quantification (Desiderio and Kai, 1983); a method commonly referred to as AQUA.

Collaboration

The overall project requires training/expertise in cell culture, sample preparation for proteomics, assays for characterizing tissue, proteomic analysis, large data analysis, statistics, programming/coding custom scripts using database management tools, etc. This doctoral project would not have progressed this far without the resources within the advisor's laboratory and collaborations in the Statistics department, and the department of Biochemistry, Molecular Biology, and Biophysics, Center for Mass Spectrometry and Proteomics, Minnesota Supercomputing Institute, and the department of Biomedical Informatics and Computational Biology. A project such as this one benefits greatly from interdisciplinary collaboration. I suggest that collaborations become more extensive in breadth and depth.

The tissue engineering researcher needs to have strength of communicating technical concepts to various disciplines. The tissue engineer(s) must communicate with proteomics experts, computer scientists, and statisticians both the major goals of the project and specific tasks at hand. Input from experts in these and other fields should be sought after continually. I recommend evaluating what the requirements of the system are, brainstorm with experts what measurements could accomplish the overall project goals and/or contain S/R data.

Bibliography

Akasaka, K., Nagahata, H., Maeno, A., and Sasaki, K. (2008). Pressure acceleration of proteolysis: A general mechanism. BIOPHYSICS *4*, 29–32.

Balestrini, J.L., and Billiar, K.L. (2006). Equibiaxial cyclic stretch stimulates fibroblasts to rapidly remodel fibrin. Journal of Biomechanics *39*, 2983–2990.

Cook, R.D. (2007). Fisher Lecture: Dimension Reduction in Regression. Statist. Sci. *22*, 1–26.

Cook, R.D., and Ni, L. (2005). Sufficient Dimension Reduction via Inverse Regression. Journal of the American Statistical Association *100*, 410–428.

Cook, R.D., and Ni, L. (2006). Using intraslice covariances for improved estimation of the central subspace in regression. Biometrika *93*, 65.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotech *26*, 1367–1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. Journal of Proteome Research *10*, 1794–1805.

Desiderio, D.M., and Kai, M. (1983). Preparation of stable isotope-incorporated peptide internal standards for field desorption mass spectrometry quantification of peptides in biologic tissue. Biological Mass Spectrometry *10*, 471–479.

Duan, N., Manning, W.G., Morris, C.N., and Newhouse, J.P. (1983). A Comparison of Alternative Models for the Demand for Medical Care. Journal of Business & Economic Statistics *1*, 115–126.

Engelmayr, G.C., Rabkin, E., Sutherland, F.W.., Schoen, F.J., Mayer, J.E., and Sacks, M.S. (2005). The independent role of cyclic flexure in the early in vitro development of an engineered heart valve tissue. Biomaterials *26*, 175–187.

Gallien, S., Duriez, E., and Domon, B. (2011). Selected reaction monitoring applied to proteomics. Journal of Mass Spectrometry *46*, 298–312.

Guo, J. (2005). Epidermal Growth Factor-induced Rapid Retinoblastoma Phosphorylation at Ser780 and Ser795 Is Mediated by ERK1/2 in Small Intestine Epithelial Cells. Journal of Biological Chemistry *280*, 35992–35998.

Hahn, M.S., McHale, M.K., Wang, E., Schmedlen, R.H., and West, J.L. (2006). Physiologic Pulsatile Flow Bioreactor Conditioning of Poly(ethylene glycol)-based Tissue Engineered Vascular Grafts. Ann Biomed Eng *35*, 190–200.

Isenberg, B.C., and Tranquillo, R.T. (2003). Long-Term Cyclic Distention Enhances the Mechanical Properties of Collagen-Based Media-Equivalents. Annals of Biomedical Engineering *31*, 937–949.

Janes, K.A., Kelly, J.R., Gaudet, S., Albeck, J.G., Sorger, P.K., and Lauffenburger, D.A. (2004a). Cue-signal-response analysis of TNF-induced apoptosis by partial least squares regression of dynamic multivariate data. Journal of Computational Biology *11*, 544–561.

Janes, K.A., Kelly, J.R., Gaudet, S., Albeck, J.G., Sorger, P.K., and Lauffenburger, D.A. (2004b). Cue-signal-response analysis of TNF-induced apoptosis by partial least squares regression of dynamic multivariate data. Journal of Computational Biology *11*, 544–561.

Jeong, S.I., Kwon, J.H., Lim, J.I., Cho, S.W., Jung, Y., Sung, W.J., Kim, S.H., Kim, Y.H., Lee, Y.M., Kim, B.S., et al. (2005). Mechano-active tissue engineering of vascular smooth muscle using pulsatile perfusion bioreactors and elastic PLCL scaffolds. Biomaterials *26*, 1405–1411.

Kessner, D., Chambers, M., Burke, R., Agus, D., and Mallick, P. (2008). ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics *24*, 2534–2536.

Kjellström, S., and Jensen, O.N. (2004). Phosphoric Acid as a Matrix Additive for MALDI MS Analysis of Phosphopeptides and Phosphoproteins. Anal. Chem. *76*, 5109–5117.

Kumar, N., Wolf-Yadlin, A., White, F.M., and Lauffenburger, D.A. (2007). Modeling HER2 Effects on Cell Behavior from Mass Spectrometry Phosphotyrosine Data. PLoS Comput Biol *3*, e4.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. Methods *25*, 402–408.

López-Ferrer, D., Petritis, K., Hixson, K.K., Heibeck, T.H., Moore, R.J., Belov, M.E., Camp, D.G., and Smith, R.D. (2008). Application of Pressurized Solvents for Ultrafast Trypsin Hydrolysis in Proteomics: Proteomics on the Fly. J. Proteome Res. *7*, 3276–3281.

Miller-Jensen, K., Janes, K.A., Brugge, J.S., and Lauffenburger, D.A. (2007). Common effector processing mediates cell-specific responses to stimuli. Nature *448*, 604–608.

Niklason, L.E. (1999). Functional Arteries Grown in Vitro. Science *284*, 489–493.

Palamakumbura, A.H., and Trackman, P.C. (2002). A Fluorometric Assay for Detection of Lysyl Oxidase Enzyme Activity in Biological Samples. Analytical Biochemistry *300*, 245–251.

Priego-Capote, F., and de Castro, L. (2007). Ultrasound-assisted digestion: A useful alternative in sample preparation. Journal of Biochemical and Biophysical Methods *70*, 299–310.

Seliktar, D., Black, R.A., Vito, R.P., and Nerem, R.M. (2000). Dynamic mechanical conditioning of collagen-gel blood vessel constructs induces remodeling in vitro. Annals of Biomedical Engineering *28*, 351–362.

Syedain, Z.H., Weinberg, J.S., and Tranquillo, R.T. (2008). Cyclic distension of fibrin-based tissue constructs: evidence of adaptation during growth of engineered connective tissue. Proceedings of the National Academy of Sciences *105*, 6537.

Thingholm, T.E., Jørgensen, T.J.D., Jensen, O.N., and Larsen, M.R. (2006). Highly selective enrichment of phosphorylated peptides using titanium dioxide. Nat Protoc *1*, 1929–1935.

Thingholm, T.E., Jensen, O.N., Robinson, P.J., and Larsen, M.R. (2008). SIMAC (Sequential Elution from IMAC), a Phosphoproteomics Strategy for the Rapid Separation of Monophosphorylated from Multiply Phosphorylated Peptides. Mol Cell Proteomics *7*, 661–671.

Turner, C.H. (1998). Three rules for bone adaptation to mechanical stimuli. Bone *23*, 399–407.

Verrecchia, F., and Mauviel, A. (2002). Transforming Growth Factor-β Signaling Through the Smad Pathway: Role in Extracellular Matrix Gene Expression and Regulation. Journal of Investigative Dermatology *118*, 211–215.

Villén, J., and Gygi, S.P. (2008). The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. Nat Protoc *3*, 1630–1638.

Weinbaum, J.S., Qi, J., and Tranquillo, R.T. (2010). Monitoring Collagen Transcription by Vascular Smooth Muscle Cells in Fibrin-Based Tissue Constructs. Tissue Engineering Part C: Methods *16*, 459–467.

Ye, J., Zhang, X., Young, C., Zhao, X., Hao, Q., Cheng, L., and Jensen, O.N. (2010). Optimized IMAC−IMAC Protocol for Phosphopeptide Recovery from Complex Biological Samples. Journal of Proteome Research *9*, 3561–3573.

Zhang, Y., Wroblewski, M., Hertz, M.I., Wendt, C.H., Cervenka, T.M., and Nelsestuen, G.L. (2006). Analysis of chronic lung transplant rejection by MALDI-TOF profiles of bronchoalveolar lavage fluid. Proteomics *6*, 1001–1010.

Zhang, Y., Ficarro, S.B., Li, S., and Marto, J.A. (2009). Optimized Orbitrap HCD for quantitative analysis of phosphopeptides. Journal of the American Society for Mass Spectrometry *20*, 1425–1434.