Genomic Response, Bioinformatics, and Mechanics of the Effects of Forces on Tissues and Wound Healing

By

Vishal Saxena

S.M., Mechanical Engineering, Massachusetts Institute of Technology, 2001

SUBMITTED TO THE DEPARTMENT OF MECHANICAL ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
IN
MECHANICAL ENGINEERING
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
MAY 2005
[June 2005]

Signature of Author: _____
Department of Mechanical Engineering
May 5, 2005

Certified by: _____
Isaac Kohane
Thesis Supervisor
Lawrence J. Henderson Associate Professor of Pediatrics, Harvard Medical School & Harvard MIT HST department, Director, Children's Hospital Informatics Program

Certified by: _____
David Wallace
Chairman, Thesis Committee
Esther and Harold E. Edgerton Associate Professor, Co-Director MIT CADLAB

Certified by: _____
Dennis Orgill
Associate Professor of Surgery, Harvard Medical School, Associate Chief, Plastic Surgery, Brigham and Women's Hospital

Certified by: _____
Lallit Anand
Chairman, Departmental Committee on Graduate Students

Genomic Response, Bioinformatics, and Mechanics of the Effects of Forces on Tissues and Wound Healing

by

VISHAL SAXENA

Submitted to the Department of Mechanical Engineering
on May 5, 2005 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosopy in
Mechanical Engineering

ABSTRACT

The mechanical environment of the cell is important for changes in behavior. Cellular behavioral changes can be traced to gene expression profile changes. These effects were studied in the context of Micromechanical force therapy, a novel therapeutic treatment in the management of different types of wounds. The mechanism of therapies that work by applying suction pressures is still not completely understood. It is proposed that micromechanical forces are the dominant mechanism by which they obtain accelerated wound healing. However, these therapies don't only impose forces to (wounded) tissue. They also remove edematous tissue as well as applies hypobaric oxygen conditions to the tissue.

Therefore, it was decided first to study only the effects of pure forces on normal tissue. 50g forces were applied to rat ears in vivo. The ears were then sampled over a period of time for their gene expression profiles on Affymetrix RAE 230 2.0 gene chips. 8 time points were obtained for each of control and stretch conditions. One rat ear was chosen to be control (and without stretch), the other was stretched. A modified Gene set enrichment analysis (GSEA) on the expression profiles was conducted using the paired difference t statistic squared as the ranking metric. A further refinement which didn't use the sign of the t statistic looked at expression changes in either direction (the combined analysis). Important genesets were obtained relevant to the differences seen between control and stretch conditions.

In the combined GSEA analysis, hypoxia came to the top of the gene set list followed by 'response to mechanical stimulus'. To test the significance of this result, a permutation test was conducted on the dataset. By shuffling the class (control versus stretch) labels randomly, different permutations were created and the GSEA was run to see if the occurrence was a chance event. The P values obtained were 0.001 and 0.014 showing that both were significant because they were not enriched randomly.

The hypoxia geneset could be an important modulator of forces in wound healing. Future work will test in a biological model the validity of the importance of hypoxia in this system.

Thesis Supervisor: Prof. Isaac Kohane
Lawrence J. Henderson Associate Professor of Pediatrics, Harvard Medical School & Harvard MIT HST department, Director, Children's Hospital Informatics Program

Table of Contents

Chapter 1

Background and Introduction

This thesis seeks to understand the effects of forces on tissues and on wound healing, and it deals primarily with effects and conditions that are of a mechanical nature. Therefore, we will define this word. Some definitions that capture the essence of our usage are the following (from Webster's online):

caused by, resulting from, or relating to a process that involves a purely physical as opposed to a chemical change.

relating to, governed by, or in accordance with the principles of mechanics. [Mechanics is...*a branch of physical science that deals with energy and forces and their effect on bodies*]

In summary, mechanical deals with matter, and in this thesis we will use mechanical rather broadly to imply "non-chemical". Mechanical energy is energy that matter has by virtue of its position or its motion. Before we define our usage of mechanical forces, we need to define forces and their types. A force is an influence that tends to change the state of rest or motion of a body. There are two types of forces: Body forces (such as gravity), and surface forces (such as contact forces between two bodies). We will mostly deal with surface forces. We will define "mechanical forces" as forces caused by hydrostatic or dynamic pressure (positive or negative), and forces caused by contact between two physical entities (such as the skin on the hand and a doorknob).

A mechanical stimuli is thus a stimuli that is a fluidic pressure (or contact pressure) or contact between two entities. Stress and strain can be considered to represent the mechanical state of a body. They are related to each other through the intrinsic mechanical properties of the body (such as Elasticity modulus, Poisson's ratio).

Stress = E*Strain

As mentioned, we seek to understand the effects of forces on tissues and on wound healing. To this end, we develop a progressively complex series of studies that seek to show the effects of forces on tissues and wound healing.

Effects of forces

Why is it important, however, to understand effects of forces on tissues? It is well known that cells respond to their chemical environment. For example, different autocrine, paracrine, and endocrine molecules travel to the wound bed to effect the wound healing response. This response is seen in the change in the gene expression profile of cells that reside in the wound bed. However, this is not the end of the story. The mechanical environment of the cell (comprising both the mechanical state of the cell surroundings and mechanical stimuli to the cell) also causes a change in the gene expression profile of resident wound cells.

For example, Lott-Crumpler [1] says that wounds heal better if their principal stresses are minimized. And, in a recently submitted paper, Saxena et al have shown that the Vacuum assisted closure device (or the VAC sponge)—we'll talk about the VAC device further down in this chapter—works by optimizing the level of strain that wound cells are placed under [2]. This device itself is one of a class of devices that we call a micromechanical force therapy device (and henceforth we shall discuss the general MMF therapy devices). A possible explanation that takes into account both these ideas could be that stresses should be at an optimal level and not too high or too low.

Now that we've rationalized the importance of studying forces on tissues, we need to decide on a model system which we can use to look at these effects. A literature search was therefore conducted to see what types of models exist on forces and wound healing. We describe these next under 'literature search'.

Our lab has looked at the rat ear as a suitable model for the application of forces to tissues. We decided to use this model for the forces on tissues study.

Importance of wounds

Worldwide, wounds pose a major health issue. Lower extremity ulcers alone cost the US Medicare system $1.5 billion. Thus, unless wound therapies see a large improvement, we will see escalating treatment costs as the population ages and as the incidence of diabetes (which itself increases with the aging population) increases.

Importance of forces in the context of wounds

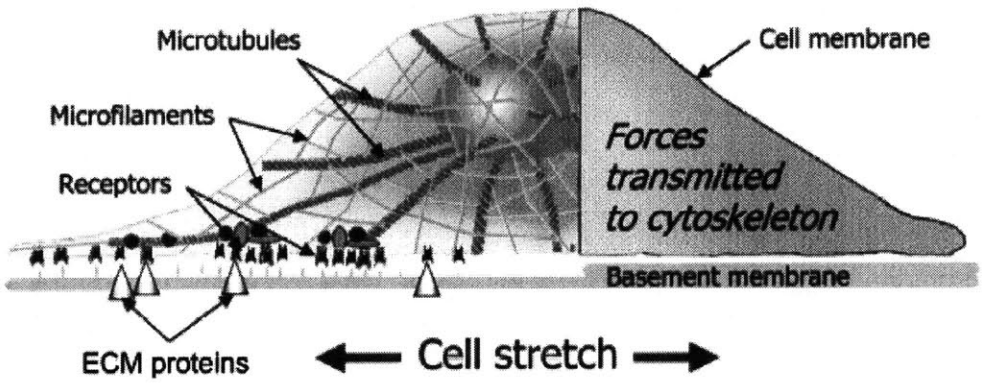
Cell proliferation and division are induced by the action of micromechanical forces. Tissue expansion is used by plastic surgeons to expand soft-tissue envelopes in reconstructive surgery, and distraction osteogenesis is used by orthopedic and maxillofacial surgeons (based on Ilizarov's work) to increase the size of bones [3, 4].

Further, during morphogenesis, cell proliferation occurs in a site-specific manner in response to local strain from tissue expansion, as happens in bud formation of glandular branches or gyri formation in the brain. These proliferative changes help to sculpt tissue architecture and shape organs [5]. Ingber et al. have shown that in order for cells to respond to soluble mitogenic factors and proliferate, they have to extend, generating isometric tension either by adherence to a stiff substrate or by external application of mechanical forces. Cells that aren't able to extend assume a more spherical shape (e.g., by spatial restriction of cell spreading, or by growing on a malleable substrate that cannotresist cell contractility and thus dissipates cell tension) become growth-arrested and tend to become apoptotic [6], [7], [8].

Thus, cells recognize mechanical signals and in response increase their rates of proliferation. We can think of this behavior as providing a natural mechanism for tissue

homeostasis. Where tissue mass expands, cells are stretched and thus divide in response. Further, our results suggest that the application of the MMF device exerts micromechanical forces on individual cells in the wound bed, thus stimulating cell proliferation and accelerating wound healing.

In previous in vitro work, it has been shown by Ingber and others that only cells allowed to stretch can divide and proliferate in response to soluble growth factors (see Fig below), whereas cells that are not stretched and assume a more spherical shape are cell-cycle arrested and tend to undergo apoptosis [7], [6]. Directional growth of capillary sprouts is also promoted by tension application in three dimensional angiogenesis models in vitro [9]. Moreover, it is known that vascular endothelial cells express a different array of genes depending on whether they were exposed to static, laminar, or turbulent flow [10], [11], [12]. The figure below is reproduced from [2].



Schematic of a cell. The cytoskeleton is attached to a basement membrane. Cell stretch is effected by attachment to the extracell matrix or by application of external force. Gene pathways are hypothesized to be triggered by microfilament/microtubule deformations resulting from cell stretch.

It is apparent that cells are able to sense mechanical forces and respond through the regulation of specific genes and the induction of cellular programs. The exact mechanisms for these effects are not fully understood but likely are related to conformational changes in the cytoskeleton in response to mechanical forces. We hypothesize that the application of mechanical force to wounds induces tissue deformation at the level of individual cells, leading to cell stretch, thereby providing a

powerful mechanism for inducing cell proliferation and angiogenesis and hence promoting wound healing.

We developed a finite element model to study the strains in wounded tissue. We found that the simulated tissue under MMF therapy conditions had strains in the range of 5 to 20 percent depending on the parametrizations that were conducted (these mimicking the stage of wound healing) that were similar to those seen in different tissues under optimal conditions for growth [2], [6], [7]. This study is described in detail further down in the thesis.

Other non-mechanical effects such as soluble growth factors and attachment to cells to extracellular matrix proteins, even though essential, are not sufficient to stimulate cell proliferation [5]. Progression of cell through its cell cycle also requires the appropriate physical context to respond to these two chemical stimuli.

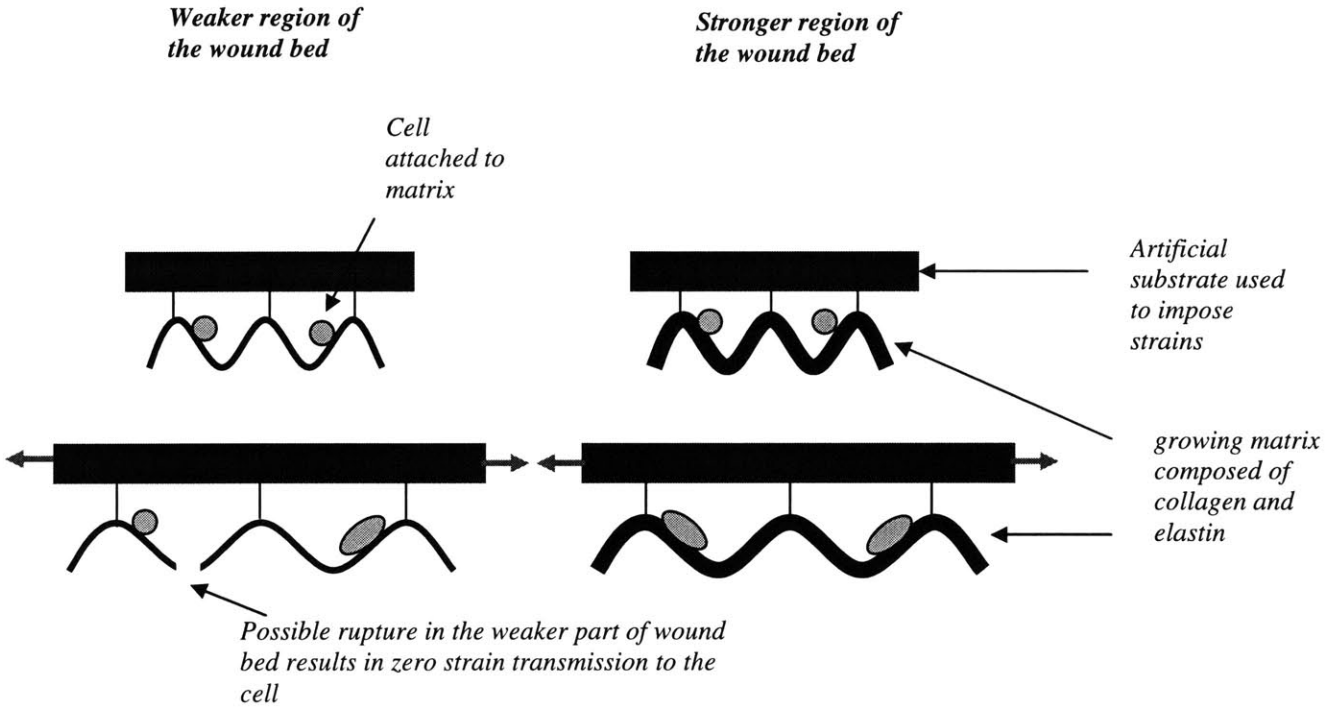
Other researchers have also shown that tissues need to be placed under specific strain fields in order to regenerate. For example, Miyajima [13] has shown that osteoblasts seeded in vitro show maximal growth at 17% strains, while Akhyari et al have shown [14] that a 20% strain (or stretch) "enhanced the formation of a three-dimensional tissue-engineered cardiac graft by improving the proliferation and distribution of seeded human heart cells and by stimulation organized matrix formation resulting in an order of magnitude increase in the mechanical strength of the graft."

However, this structural requirement may be absent in wounds, as the normal matrix is degraded or altered such that it fails to form a mechanically continuous structure to provide the scaffold on which cells normally stretch and build up isometric tension (for example as may happen by stretching the scaffold through the remaining intact network and also as may happen when cells become seeded onto the MMF therapy 'sponge' material)—see the section below on forces are transmitted through the wound bed and the Figure in the section.

The mechanical strain generated by the MMF therapy in the microscopic tissue domains may, in addition to other benefits of suction, overcome this loss of tissue integrity and substitute for the missing structural basis necessary for cell proliferation.

A conceptualization of force-transmission to cells in the wound bed

Forces applied at boundary conditions always (unless there is rupture in the tissue) translate to constant forces in all areas of the wound bed (at steady state conditions).



Optimally, we'd like to impose a constant strain at each micro loci, as it is optimal strains that are critical for cells to proliferate and divide. These optimal strains do not necessarily have to be constant across different areas of the wound bed—they may for example be lower for regions that have a lower modulus of elasticity or weaker strength. In other words, the tissue will display non-homogeneous behavior in terms of material properties. This could for example be because if the tissue at the weaker spots is less elastic, it may rupture bringing the strains to zero, even though the applied strains through a scaffold may seem to be non-zero.

When an external global force is input, the force has to be lower than the yield stress of the weakest area of the wound bed. Otherwise the force will cause that area to expand until it breaks, and that will stop any further force transmission through that axis.

It may also happen that the 'lodging' of the ruptured or wounded tissue within the 'lattice' network of the sponge may facilitate the transmission of forces to tissue (and therefore to cells) that makes up the wound bed.

Thus the MMF therapy sponge—which we'll talk about next—facilitates force transmission by in essence providing a scaffold onto which the (at least the) surface areas of the wound bed can 'lodge' on and which then when subjected to a negative pressure are 'pulled' inward into the sponge—the sponge providing contact points where resistance 'supports' the wound surface creating continuities that may possibly have been broken because of the rupture in the intact tissue scaffold.

(One of) the model systems that we used to study these force effects is Micromechanical force therapy (MMF therapy)

A device that is now commonly used in wound closure clinical applications is the Vacuum Assisted Closure Therapy (VAC) device (KCI, San Antonio, Texas)—one device in the general class of MMF therapy devices. The VAC is a relatively new therapeutic paradigm. Its efficacy has been well documented [15], [16], [17]. For example, according to [15], the VAC obtained a 66% change in depth versus 20% for "saline-wet-to-moist dressing". Because the VAC imposes (we'll talk about the VAC in more detail shortly) negative pressures on the wound bed, it is a good model system in which to explore the action of mechanical forces on the process of wound healing (the VAC obtains impressive accelerations in the wound healing process).

However, the mode of action of MMF therapy devices is still not completely understood. Let's look at some of the theories that have been put forth, and then we'll talk about our

hypothesis—which says that micromechanical forces play a vital role in the VAC sponge's mechanism of action in accelerating wound repair.

Many theories have been advanced to explain the physiologic basis for the marked improvement in clinical outcomes achieved by the MMF device. One of these proposes that the application of suction evacuates interstitial fluid and cellular debris and reduces local edema, decreasing the likelihood of wound infection [18]. Although this may be an important mechanism for a selected subset of wounds, we have observed many MMF therapy-treated wounds in which minimal fluid was extracted; nevertheless, in these wounds, dramatic healing responses were observed.

Also, it is proposed by many that the MMF device creates a hypobaric pressure which then causes an increase in blood flow to the wound bed. These effects have been shown to accelerate the formation of granulation tissue [19], [20], [21], [22]. It is interesting to note that intermittently applied subatmospheric pressure has produced superior results, possibly because of mitigating the cellular desensitization that occurs with exposure to continuous subatmospheric pressure [18].

In addition, many have noticed increased granulation tissue, decreased bacterial levels, and increased cell growth in MMF-treated wounds. Dr. Orgill and colleagues at the Brigham and Women's Hospital have observed similar changes in their own patients. However, we believe that the cause may be secondary to an underlying biologic effect.

Thus, although it is likely that all of the above factors play a role in the action of the MMF device, we propose that the application of micromechanical forces to the wound site may be the most significant mechanism of action.

For example, directional growth of capillary sprouts is also promoted by tension application in three dimensional angiogenesis models in vitro [9]. Moreover, it is known that endothelial cells in the vasculature express a different array of genes depending on whether they were exposed to static, laminar, or turbulent flow [10], [11], [12]. It seems

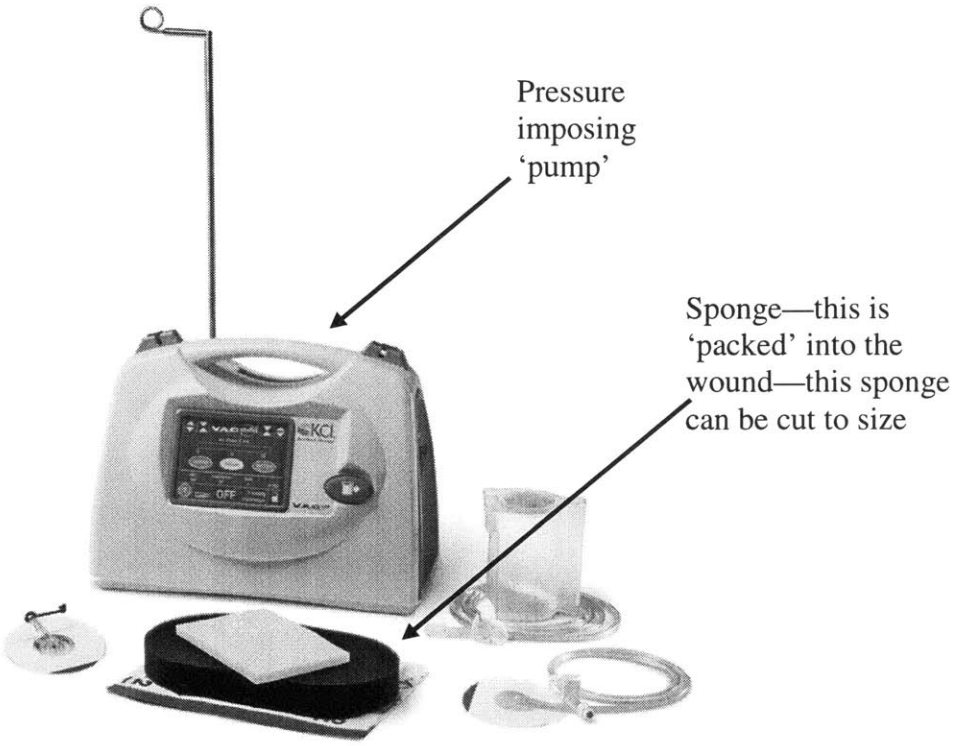
that cells are able to sense mechanical forces and respond through the regulation of specific genes and the induction of cellular programs. The exact mechanisms for these effects are not as yet completely understood but likely are related to conformational changes in the cytoskeleton in response to mechanical forces. It is our hypothesize that the application of mechanical force to wounds induces tissue deformation at the level of individual cells, leading to cell stretch, thereby providing a powerful mechanism for inducing cell proliferation and angiogenesis and hence promoting wound healing

At the Brigham and Womens Hospital, the MMF device has become the most common treatment modality for complex wounds. The device consists of a porous open-cell sponge (polyurethane or polyvinyl alcohol) applied to a wound and covered with an occlusive dressing. The sponge has a tube connected to a vacuum pump which applies subatmospheric pressure to the sponge (see Fig below). In a porcine wound model, the MMF device was shown to increase blood supply and reduce the incidence of infection [19]. The MMF device has been proven to be effective at treating both acute and chronic wounds [20], [21], [16]. Thus, with the above background, we wanted to investigate other possible mechanisms of action that may help explain how the MMF device works.

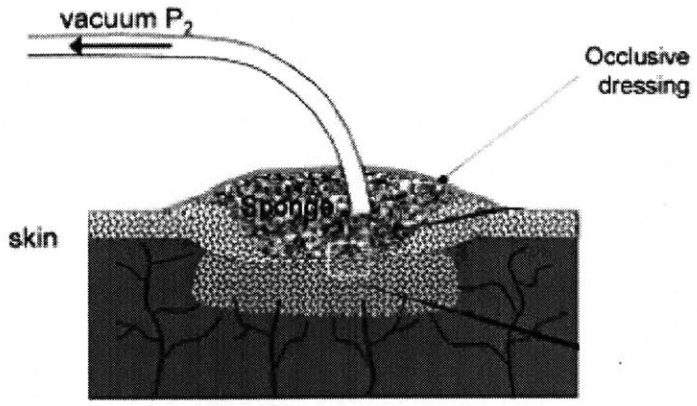
Let's quickly look at one example of the MMF device. A picture of the VAC system is shown next [23]:

Another product from KCI is shown next [24]. This product "...combines VAC therapy with Instillation therapy to assist in the irrigation and cleansing of wounds and the removal of infectious materials" [24].

The VAC works by imposing a negative pressure on the surface and margins of wounds. The device applyies this pressure through a special dressing placed in the wound cavity or over a flap or graft [24]. Currently, the therapy is indicated for "chronic open wounds, diabetic ulcers, pressure ulcers, acute and traumatic wounds, flaps and grafts, dehisced wounds, and partial thickness burns" [24].



A schematic of a MMF therapy device is shown next [2].



Given that the MMF device works, we'd still like to know why it is important to study the effects of forces on tissues with *this* particular therapeutic paradigm. What makes the MMF therapy so pertinent is that it is one of the few effective therapies that can treat the types of wounds described above (it is also quite suitable as an adjunct therapy, being used to treat wounds in the acute time frame after which other therapies can take over) [25]. The market for such wounds is large. Let's take a quick look at the market forces driving this product.

MMF market analysis

According to [25] the total market for the VAC was estimated to be 1.4 million patients in the US in 2004 with an approximate growth rate of 13%. KCI obtained revenues of $992.6 million in 2004 and this is projected to rise to $1.5168 billion by 2006. Thus, the VAC is an important therapy [25].

The VAC applied to human patients however forms the tail end of our analysis. To better understand the effects of forces on wound healing, we need to start by decoupling the effects of forces on wounds and the effects of forces on tissues. By studying in isolation the effects of forces on tissues, we can then do the more detailed analysis taking wounds into account.

The strains imposed by the MMF device are significant

Further, before we even start, we need to make sure that the wound does indeed see forces and stresses that are strong enough to impact the tissue physiologically. However, what defines physiologic strains in the wound (and by extension physiologic stresses)? To answer this question, we looked at other wound and tissue systems where the strains have more clearly been measured.

For example, Miyajima [13] has shown that osteoblasts seeded in vitro show maximal growth at 17% strains, while Akhyari et al have shown [14] that a 20% strain (or stretch) "enhanced the formation of a three-dimensional tissue-engineered cardiac graft by improving the proliferation and distribution of seeded human heart cells and by stimulation organized matrix formation resulting in an order of magnitude increase in the mechanical strength of the graft."

We then developed a finite element model to study the strains in wounded tissue. We found that the simulated tissue under MMF device conditions had strains that were similar to those seen in different tissues under optimal conditions for growth [2]. This study is described in detail further down in the thesis.

Decoupling force study from force study in wounds—gene expression study of rat ear stretch

After validating that the strains set up in the wounded tissue exposed to the MMF device are reasonably high enough to cause physiologic responses, we then designed an animal model where we could study the effects of forces on tissue. Prior work done at our lab has studied the rat ear because it is very thin and can be easily visualized under the light microscope. This makes studying the vessels that drain it possible. We extended this model for our tissue response to force study.

Possibly the best way to characterize the response of a tissue to a stimulus is to look at the changes that occur in the network of protein interactions that make up the cellular machinery. However, current technology doesn't allow us to do this globally within the cell. A next best way to characterize the tissue response is to study the gene expression changes that occur with a stimulus. This is therefore what we did next. Later down in the thesis, we will discuss the rat model that was developed as well as the gene expression study that was conducted on the tissue obtained from this model.

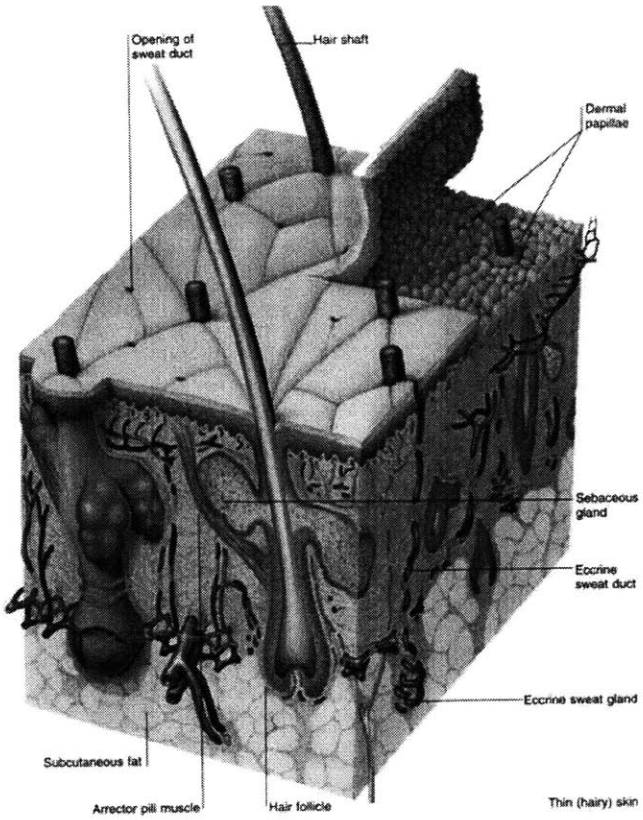
Since both our models of analyzing forces (rat ear and human skin) both deal with skin, we need to understand the structure of skin. We do this next. After this, we need to understand the series of events that take place once skin is wounded. We discuss that after we look at normal skin.

Brief review of the structure of skin (epidermis and dermis)

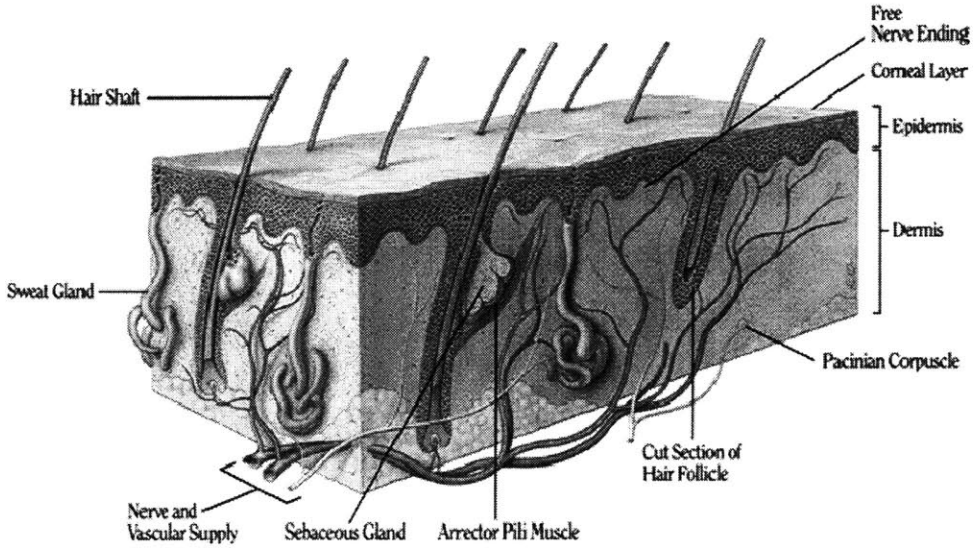
A cross section of the skin is given in the figure below. "The epidermis is a stratified squamous epithelium that mainly serves as a protective barrier. The epidermis is about 0.1 mm thick, but on the palms and soles, the thickness can be greater (0.8-1.4 mm). The keratinocyte is the principal cell of the epidermis and it serves to produce the protein keratin" [26].

The thickness of the dermis can range from 0.6mm in the eyelids and 3mm or more on the back, palms, and soles. The dermis can be divided into a thin upper layer, called the papillary dermis, beneath the epidermis interdigitating with the epidermal rete ridges. The papillary dermis is composed of loosely interwoven collagen. Deeper down is the reticular dermis, composed of coarser and horizontally running bundles of collagen [26]

Collagen fibers constitute 70% of the dermis. They impart structural toughness and strength. Elasticity in the skin comes from Elastin fibers that are arranged in all directions in the dermis. Elastin is found more near the hair follicles and sweat glands (less in the papillary dermis) [26]. The following figure is from [26].

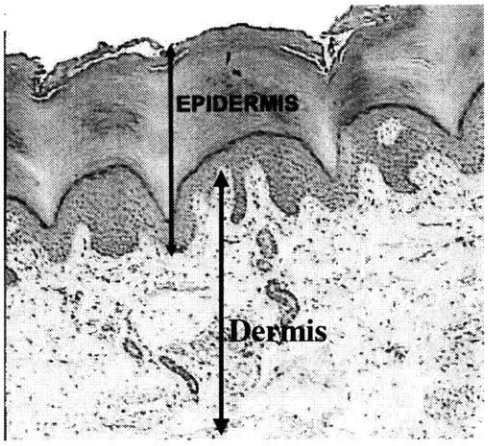


Here's another cross section of the skin [27].



"The dermal ground substance consists of a semi-solid matrix of glycosaminoglycans (GAG) which impart movement to some dermal structures. In addition, the dermis

contains: fibroblasts which synthesize collagen, elastin, other connective tissue and GAG; dermal dendrocytes which are dendritic cells with immune function; mast cells; macrophages; and lymphocytes." [26]. The following is also from [26].



Wound repair—concepts and events

Now that we've talked about the normal cross section of the skin, we will talk next about what happens during the process of wound repair. This will lay the background upon which we will impose mechanical force effects.

There are 4 phases in wound repair: a) hemostatic phase in which blood loss is arrested, b) inflammatory phase, in which inflammatory cells recruited to the wound site clear up the debris and control infection, c) proliferative phase, in which granulation tissue and a provisional matrix are laid down and reepithelialization occurs, and d) remodeling phase, in which matrix is remodeled and wound is contracted [28].

In the hemostatic phase, the blood loss is arrested through the formation of a platelet plug enmeshed in a network of insoluble fibrin molecules [29].

In the inflammatory phase, neutrophils and macrophages are the two cell types recruited. Although neutrophils play largely a microbicidal role, wound healing is severly diminished in the absence of macrophages, whose secretions include cytokines that

attract and influence the next cell types involved in wound healing (fibroblasts, endothelial cells, and epithelial cells).

The proliferative phase is the next phase. It only begins after inflammation subsides (a chronic lesion occurs if inflammation doesn't subside). This phase is defined by the formation of granulation tissue (which is loose connective tissue with a high level of vascularization) [29], and epithelialization.

During this phase, the wound facing side of the basement membrane [29] is degraded by the endothelial cells that migrate along fibronectin and other matrix components [30]. The new endothelial cell junctions are leaky. Macrophages recruit Fibroblasts which then lay down the matrix framework of collagen III and I, gycosaminoglycans, and proteoglycans [30].

Wound contraction, the final phase, occurs through either of two proposed mechanisms: 1) matrix remodeling and reorganization, and b) myofibroblast activity [29].

Rest of the thesis overview

We start out first, in the next chapter, by studying the different models of the effects of forces on tissue and different models of wound healing. We then end by explaining why it is we decided to use the three models that we used (finite element mathematical model, rat ear model, VAC device on human model). The chapter after the next, we then continue our study by exploring the forces seen in a simulated wounded tissue and the scale of these forces. This will tell us if the forces are large enough to create any meaningful effects. The surface strains are then determined to see if they are in the range of strains seen in other tissue types.

Chapter 2

Literature review: wound healing models, study of forces on tissues models

Wound healing: models

Any given wound is a complex and unique occurrence whose state depends on the unique circumstances that created it. In order to be better able to understand [31] and classify different wounds (and also to pinpoint specific expression profiles in relation to isolated stimuli and wound states), we need to look at different *reproducible* models that are currently in use or have been used for wound healing.

Experimental models are also useful for evaluating safety and efficacy of therapies for wound healing [31]. There are essentially three types of models (4 if we count mathematical models). a) in vivo animal models [31], b) in vivo human models, c) in vitro models [31] [32], d) mathematical models [2]. In vitro models can themselves be divided into singlecell systems, multicellular systems, or organ cultures [31].

The animal models can further be classified under type of wound created. Thus, [32] talks about (in turn): Excisional wound healing; reepithelialization; incisional wound healing etc.

Many animal models exist to represent the state of healing of wounds in humans [33]. The *validity* of a model rests on its closeness in being able to represent this state faithfully. Lindblast [33] says that "...we are not in need of models per se, but rather models that more accurately reflect the biological processes occurring in humans during normal wound healing." For example, if a model is representative of the wound healing in humans, then the cells will go through a similar evolutionary process, secreting similar growth factors and cytokines. Other metrics relating to the healing process will also tend

to be similar. For example, the rate of wound closure should also be the same, as should also be histologic and immunohistologic sequence of events. Other metrics include optical-visual techniques, ultrasound, and computer imaging [34].

One model discussed in [34] uses pigs to represent human epidermal wound healing. According to this paper, pigs and humans share many commonalities in their wound healing process as well in the anatomy of their skin. The relative thicknesses of the dermis and epidermis are similar in both; both have epidermal reteridges (which are the intrusions between the dermal papillae—that are the ridgelike lines in the hand and foot; in other words, the reteridges define the valleys between the protruding ridges on the hands and feet). There are other similarities listed in [34]. The pig skin however differs from the human skin in having an elastic membrane in the hypodermis (a layer of cells below the epidermis).

This membrane therefore prevents any wound study from going beyond 28 days when the wound is completely closed in the pig. The authors in [34] tried to circumvent this by using a PTFE chamber that sought to arrest this wound closure through this mechanism. (these authors also subjected the wound tissue to mechanical tests to evaluate the tensile properties of the scar tissue formed).

Reproducibility in creating the wounds is a problem. For example, the authors in [32] chapter 2 (methods in reepithelialization) talk of the difficulty in creating constant thikness epithelial wounds.

The pig was not chosen as a model system because it is difficult to keep pigs because of their size. Further, since we wanted to use Affymetrix chips in our analysis, we decided to use a rat model. Affymetrix chips for Rattus Norwegicus are available. Further, previous work at our lab has looked at effects of stretch on rat ears.

Murray [35] attempts to create mathematical models of epidermal and dermal wound healing, saying that although wound healing is a complex process that is badly

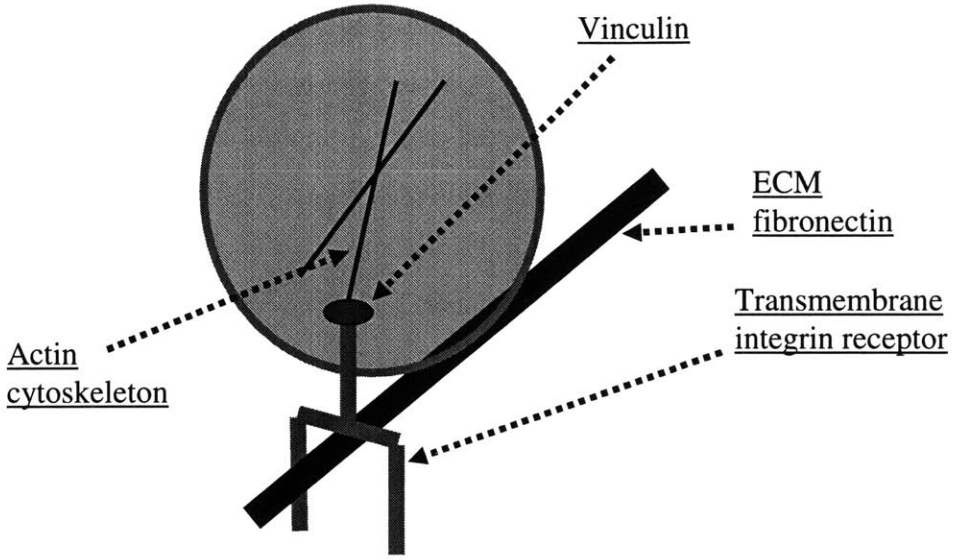
understood, only mathematical models can convert "an understanding of the underlying mechanisms into a predictive science. Another attempt to model wound healing[35] is performed by Sherratt. *In keeping with this idea of using quantitative tools in modeling wound tissue behavior, we used finite element analysis in obtaining strain and stress profiles of simulated wound tissue.*

Wound healing: mechanical effects

Cells are able to sense various types of stimuli, and, "...are known to orient and migrate in response to gradients of light intensity, electrostatic potential, and gravitational potential" [30]. Further, "...transient mechanical stimuli can induce motility of stationary fish epidermal keratocytes" [30], and mechanical tension can induce sensory and brain axons of chick neurons to be elongated [30]. Cell substrate rigidity also affects their motility and their morphology [30]. "Directional cell movement can be promoted by direct mechanical distortion of cells, for example, by pushing or applying fluid pressure with a pipette [30].

Wound healing: mechanical models

An in-vitro model by Ingber et al in [30] showed that cells take on the shape of a micro-islands over which they are placed. Further, in the presence of motility factors such as PDGF cells extend their lamellipodia preferentially from the corners onto non-adhesive areas outside the islands [30]. Cells contain motility receptors that are stimulated by PDGF (platelet derived growth factor) ([30]—page 1199). These motility receptors do not localize to corners when cells are grown on square islands [30]. However, vinculin molecules do localize to the corners as do stress fibers of actin [30]. The following shows pictorially what is stated in [30] page 1199.



The paper concludes that there is "tension molding" of the actin network (page 1202) because the cell is distorted into a square shape. This causes lamellipodia to be focused at the corners (9). "The novel finding here is that this targeting of isometric tension to the corners of square cells somehow creates a localized microcompartment in which formation of lamellipodia, filopodia, and microspikes is promoted in the absence of a gradient of soluble chemoattractants." [30].

Wang [36] tested whether integrins connect directly to the cytoskeletal framework in cells or whether the integrins connect to the outside membrane which then connects to the network within cells. To test this, a strain was applied to the integrin receptor. If this strain was supported by the cytoskeletal network directly, then the disruption of the network would release the force of resistance to strain on the receptor. Thus, when cytochalasin was input into the system, there was an increase in angular strain verifying that the integrin receptor is coupled to the cytoskeletal network directly (although why the author mentions "for only 15 minutes" on page 1125 isn't clear to me). Wakatsuki [37] has shown that cytochalasin D affects the mechanical properties of cells (cytochalasin, disrupts the cytoskeletal network in cells).

Dembo [38] shows that the traction forces used to move fibroblasts forward are applied by the lamellipodia [at the focal adhesions], while the rest of the cell body is passively pulled forward. Traction is also studied by Wang [39] who shows that if cells take on the shape of a square island, the traction is highest at edges.

Waters et al [40], state that mechanical forces involved in ventilator-induced lung injury may cause an elevation of proinflammatory cytokines, and that "microarray analysis of human lung epithelial cells demonstrated that cyclic mechanical stretch alone profoundly affects gene expression.

Mechanical forces gene expression (in fibroblasts and other cells)

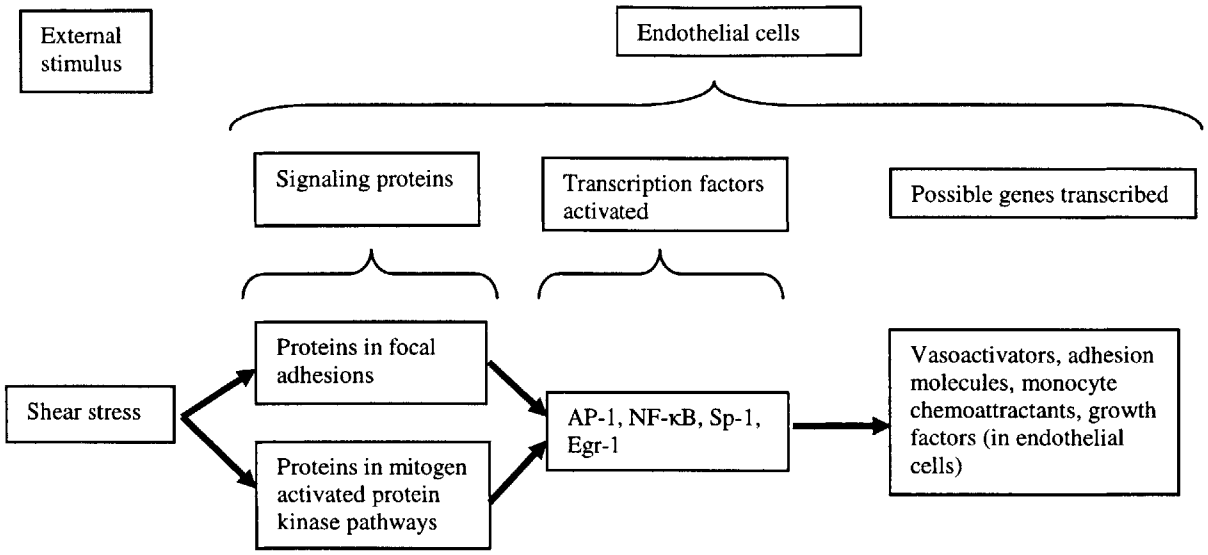
The extracellular matrix distributes mechanical forces applied to it to adhesive structures in the cell (including integrins) [41]. These forces then are transduced into intracell signals that rearrange cytoskeletal proteins and cause the expression of cytoskeletal genes such as alpha-skeletal, alpha-smooth muscle actins, filamin A, talin, and vinculin. "Filamins are actin-binding proteins that organize actin filaments into orthogonal networks and enhance the rigidity of the actin cytoskeleton." [41]. Thus the force that is applied to the cell causes the cell to essentially become more stiff thereby resisting the deforming effects of the externally applied force. This paper [41] says that forces cause disruption of the microtubule structure in the cells leading to increased filamin A which causes an increase in actin levels thus causing increased rigidity. Compounds that also disrupt the microtubule assembly also lead to increased actin through the increase in filamin transcription (if the microtubule is stabilized, the actin assembly is not affected).

Chiquet et al [42] have shown that tenascin-C (an ECM component) is directly regulated by mechanical stress. "Fibroblasts sense force-induced deformations (strains) in their ECM," and forces cause the fibroblasts to increase the induction of the mRNA of tanascin-C directly without the release of other factors into the medium [42]. The fibroblasts need to be pre-stressed in order to be able to induce tenascin-C in response to an external force. If they are not pre-stressed, they don't respond to forces with the

transcription of tenascin-C. The "integrins within cell-matrix adhesions can act as strain gauges, triggering MAPK and NF-κB pathways in response to changes in mechanical stress" [42].

Kraiss et al report that fluid flow through the application of shear stress can independently activate pp70W^{S6k}, which is a key intracell "checkpoint in the transduction of mitogenic and other extracell signals [43]. This paper also shows that the response that they observed (the activation of pp70W^{S6k} with fluid shear) was cell type dependent. Thus they didn't see this same response in CHO cells, or Swiss 3T3 fibroblasts [43]. Although the effects of different cell types to forces as we saw in ref [43] are cell type dependent, the principal argument is that forces can affect the transcriptional/translational paradigms in cells.

Chien et al show the effects of mechanical forces on signal transduction and gene expression in endothelial cells [44]. This paper says that shear stress is dependent on blood flow and vascular geometry, and strain is dependent on transmural pressure, the mechanical properties of the vessel. It would perhaps be more appropriate to specify that the boundary conditions impose a shear stress causing that variable to become independent and then making the strain the dependent variable (Stress and strain are coupled). The ideas in this paper are encapsulated in the following relation:



Another paper dealing with shear stresses is by Ji [45]. Low shear is suspected of being implicated in atherogenesis (low shear is found at bifurcations where the flow is characterized by vortex flows). Steroids which act at the glucocorticoid receptor are known to reduce inflammation. This paper reports that shear stresses cause activation of the glucocorticoid receptor, thus establishing the link more concretely between shear and the lower atherosclerosis seen at regions of the vasculature with comparatively higher shear stresses [45].

Because prior work at the Orgill lab has studied that rat extensively and because rat (or Rattus Norvegicus) Affymetrix chips are available, and also because rats are far easier to keep and maintain than larger animals such as pigs, we decided for the animal in vivo study to use rats in our study.

Further, the force input was applied to the rat ears mainly because the Orgill lab has had extensive experience applying forces to the rat ear. One of the reasons the rat ear has been chosen is that changes in the rat ear vasculature can be looked at directly under a light microscope (since the rat ear is thin enough to facilitate this).

The mathematical model we used in our wound study to study the mechanical effects used finite element analysis as the implementation tool.

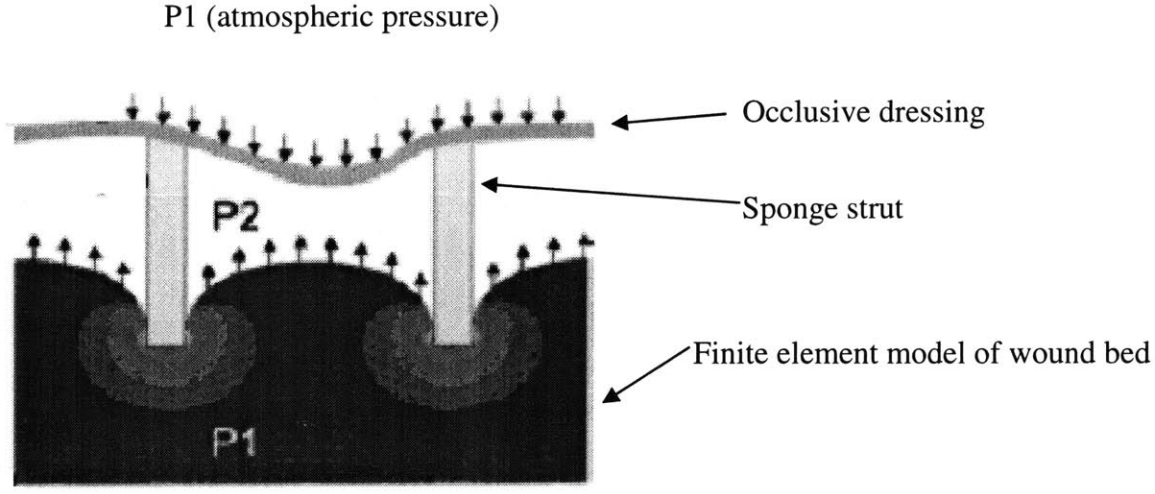
First, we study the strains that are obtained in a skin wound model applying finite elements. This study will give us insights into a mechanism of the VAC device that we feel is significant in its efficacy—namely the effects of micromechanical forces in obtaining enhanced outcomes. That the VAC obtains improved outcomes is not in doubt. Thus we have a system that is highly mechanical in its effect on the skin (or at least seems so on first inspection) and therefore is a good candidate to study the effects of forces on tissues.

Chapter 3

Vacuum-Assisted Closure: Microdeformations of Wounds and Cell Proliferation

As mentioned in the introduction, we wanted to test whether the strains obtained through the application of the VAC are significant enough. Thus, we created a finite element model of skin. A solid elastic linear, isotropic, homogenous model was constructed (thus there were no empty spaces in the model) and fluid flow through this model was not taken into account. The gage pressure to which this model of skin was exposed was set at 110mmHg (this was kept constant for all the analyses except the analyses where the pressure itself was varied).

A finite element model was constructed. Further down we'll explain the specifics of the model. The figure below shows that a mismatch between the compliance/elasticity of the VAC sponge material and the tissue causes the tissue to intrude into the spacing between the struts.



We have also compared the microdeformations of the wound bed to our finite element model. As seen in the figure below, we find that our finite element model of VAC

application to a wound bed predicting its microdeformations are comparable to histologic evidence of wound tissue microdeformation in clinical wounds treated with the VAC (see Fig below). This lends at least a somewhat qualitative validity to the finite element model.

The use of finite elements in plastic surgery applications is well established. It is used in applications ranging from the analysis of craniofacial stress, [46], to burn heat transfer, [47] to skin deformation [48].

The wound was simulated using mechanical properties from the literature and assuming that these properties were constant in all directions (*in other words we assumed the wound bed to be isotropic*). By altering the distance between struts and the thickness of the struts, we changed the physical properties of the sponge referred to as pore diameter and *pore volume fraction*. We further demonstrate how the finite element analysis model can be used to optimize the set of conditions that define the VAC system.

MATERIALS AND METHODS

Histologic Processing

Background histologic work to obtain tissue microdeformations was performed by colleagues at the Orgill lab.

Routine biopsy specimens of five clinical wounds treated with the VAC for 4 to 7 days were stored in formalin, embedded in paraffin, sectioned in 5μm sections perpendicular to the wound surface, and stained with hematoxylin and eosin. Biopsy sections were evaluated for surface undulations and cellular and vascular structures.

Finite Element Modeling

ADINA Version 8.0 (Adina R&D, Watertown, Mass.), finite element software, was used to formulate the finite element analysis simulations in this study. The validity of these

analyses was checked by the solution of closed form analytic equations. An order of magnitude estimation is included in Appendix (put appendix number here).

The VAC Computational Model

The VAC device consists of a highly porous polyurethane sponge with spatial connections between the pores, allowing subatmospheric pressure applied to the sponge to be distributed equally throughout the sponge. The VAC device used today applies a 70- to 150-mmHg vacuum. The VAC is modeled as a series of pores with the polyurethane sponge in contact with the wound. We assume the pores to be symmetric and that there is no lateral displacement in the center of the pore. Furthermore, where the wound contacts the sponge, we expect no vertical displacement. These *boundary conditions* and pressure are applied in a two-dimensional finite element model. The wound is modeled as a *linear*, homogeneous, isotropic, *elastic* material. Although a nonlinear stress-strain relationship exists for the skin and other connective tissues, the skin shows a fairly linear stress-strain curve in the observed strain ranges [49]. Furthermore, we varied the compressibility of the material (*Poisson's ratio*) to study this effect on wounds. Five parameters were studied: stiffness of the wound (*Young's modulus of elasticity*); compressibility of the wound (Poisson's ratio); pore diameter of the sponge, defined by the distance between two struts in the model; pore volume fraction of the sponge; and pressure applied to the sponge (or differential thereof). We modeled variation in each of the above parameters. The figure and Table below are from [2]. As each parameter was sequentially varied, all other parameters were assigned "standard" values usually observed in the application of the VAC. See Table below (from [2]).

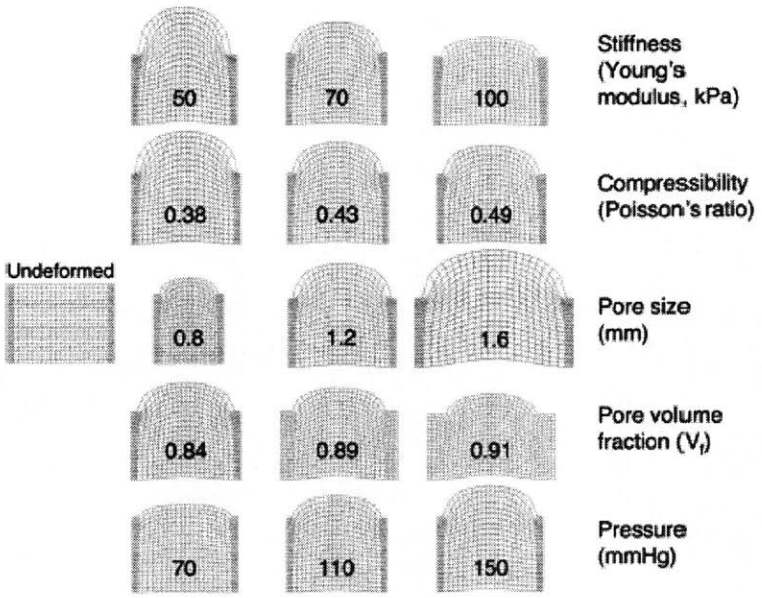


Table: Finite element parameters studied

	E (elastic modulus, kPa)	*L* (pore diameter, mm)	*N* (Poisson's ratio)	*P* (pressure, mmHg)	V_f (pore volume fraction)
Varying E	50 70 100	1.2	0.49	110	0.889
Varying L	70	0.8 1.2 1.6	0.49	110	0.842 0.889 0.914
Varying N	70	1.2	0.38 0.43 0.49	110	0.889
Varying P	70	1.2	0.49	70 110 150	0.889
Varying V_f*	70	1.2	0.49	110	0.889 0.842 0.762

* Changes to V_f use strut sizes of 0.15, 0.3, and 0.5mm; 0.15mm is used when strut thickness is not varied.

Simulation Parameters

Constraint boundary conditions. The system is modeled as a large number of pores with associated struts such that the centerlines underneath the struts impose a symmetric boundary condition. At the centerline underneath each strut, we assume no lateral displacement. Vertical displacement will be slight because of the almost incompressible

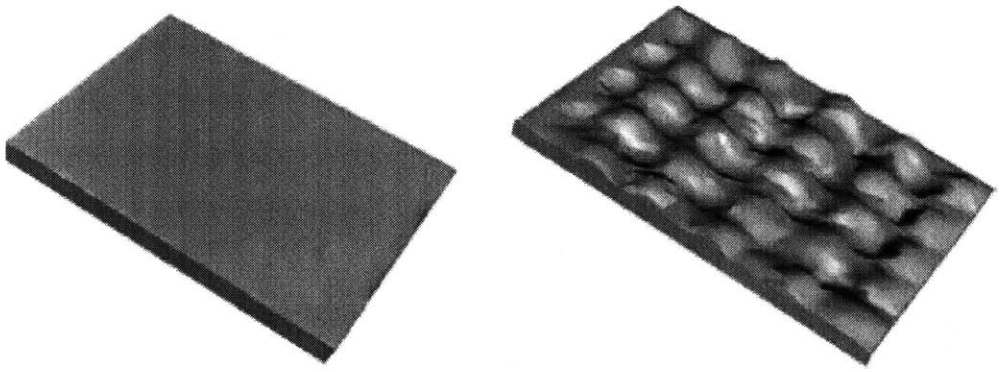
condition imposed on Poisson's ratio. This symmetric boundary condition was modeled using half the strut thickness in the analyses.

Force boundary conditions. In this static case, no shear *stresses* are imposed on the wall; therefore, all the forces act perpendicular to the wall. Thus, we impose only a uniform pressure boundary condition in our analysis.

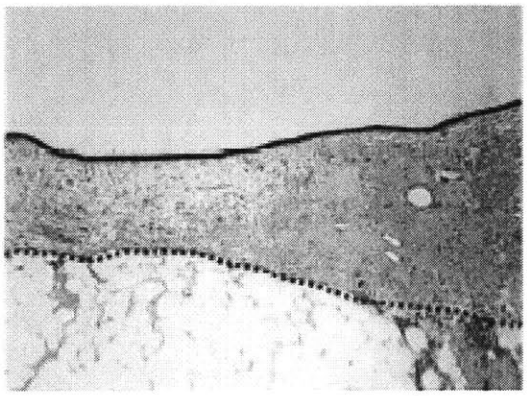
RESULTS

VAC Device Increases the Microscopic Surface Area of the Wound

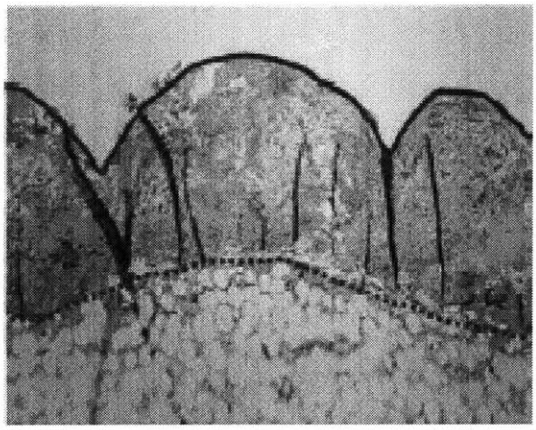
Histologic sections of wounds were studied comparing areas treated with a VAC for 4 to 7 days to other areas in the same wounds not in contact with the VAC sponge. VAC-treated wounds show a marked increase in undulating contour with protrusions and indentations corresponding to the geometry of the sponge's contact with the wound (see Figure below—from [2]).



Areas underneath an occlusive dressing only, without sponge contact, do not undulate (see Figure below—from [2]).



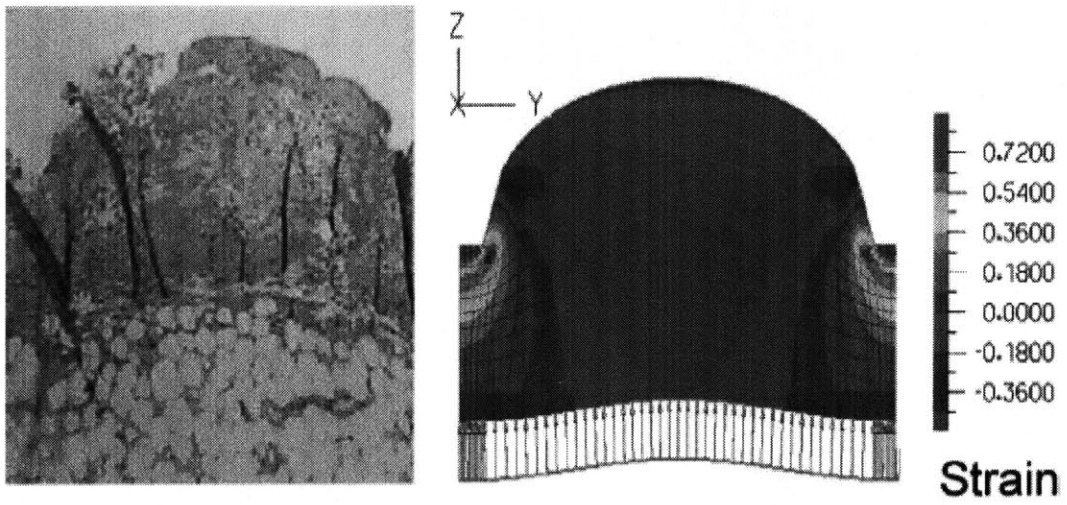
The following figure shows how areas under the VAC applied suction with the sponge show microundulations (from [2]).



These induced surface irregularities cause an increase in microscopic surface area, and thus local mechanical distention of tissue, without increasing the actual size of the wound. Measurements of the surface length of a histologic cross-section of a wound demonstrate an increase in surface length of 22 percent over 4 days compared with wounds not treated with the sponge. In addition, a rich vascular network was noted to be present in the wounds treated with the VAC sponge compared with control sites (Figure shown above).

Finite element analysis of the wound-VAC system was conducted to directly predict the *strain* imposed by the VAC sponge on wound tissue. Visible in the finite element

analysis is tissue stiffness (characterized by Young's modulus), tissue compressibility (characterized by Poisson's ratio), sponge pore diameter, pore volume fraction, and variability of wound tissue strain over a range of imposed pressure (Fig.
A above). Higher strains can be induced by increasing pressure, increasing pore diameter, or decreasing strut thickness. Strains are also predicted to be greater when mechanical properties of the wound such as the Young's modulus or Poisson's ratio are decreased. Using physiologic values for tissue, [49], [50], typical sponge pore diameter, and pressures used clinically results in a striking resemblance between the finite element analysis output and the histologic cross-section (Figure shown below—from [2]).

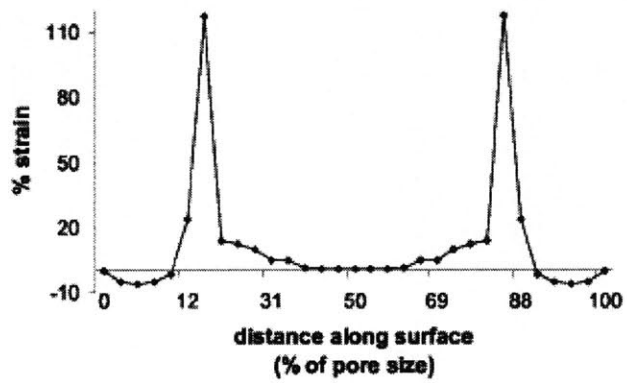


Point-wise strain along the simulated wound surface was maximal in the regions close to the struts and, consistent with a model of a thin membrane acted on by a uniform pressure, nearly constant across the majority of the wound tissue.

Strain Variability along the Wound Surface

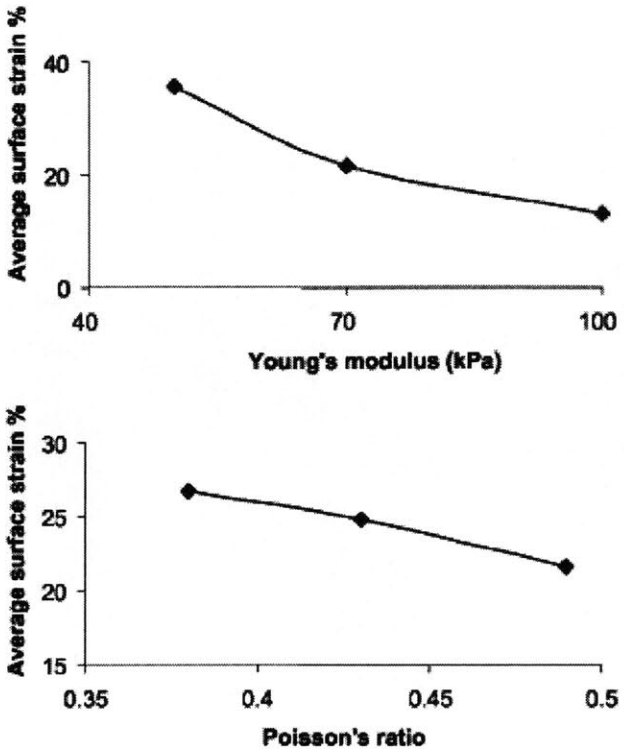
Surface strain varies in a repeating pattern across the wound tissue. Strain is negative (compression) immediately underneath the sponge struts, as they indent the surface of the wound. At an applied pressure of 15 kPa (110 mmHg) typical of the VAC, the tissue quickly reaches a peak strain of 125 percent immediately adjacent to the struts (0.15 mm), as the forces of strut compression and vacuum suction oppose each other

maximally and most directly at the edges of the sponge pores. The bulk of the wound tissue within the sponge pore, however, experiences lower strains (5 to 20 percent), with tissue at the center of the sponge pore experiencing the lowest strains (up to 5 percent) (see Fig below). These strains are dependent on wound thickness. At a wound thickness of 1 mm, the center strains were 0.67 percent, whereas for the most superficial wounds (0.5 mm), the center strains were 5.1 percent. The figure below is from [2].



Wound Healing Affects Tissue Strain

During wound healing, tissue elasticity and compressibility change, [49] and wound displacements induced by the VAC, even at constant pressures, are likely highly time-dependent. As wounds heal, they tend to become fibrotic with increasing stiffness (Young's modulus of elasticity), thereby decreasing average wound strain. For example, increasing the stiffness from 50 kPa to 70 kPa would lower the average wound surface strain from 35 percent to 12 percent, with a concomitant decrease in peak strains from 200 percent to 125 percent (see figure below—from [2]).

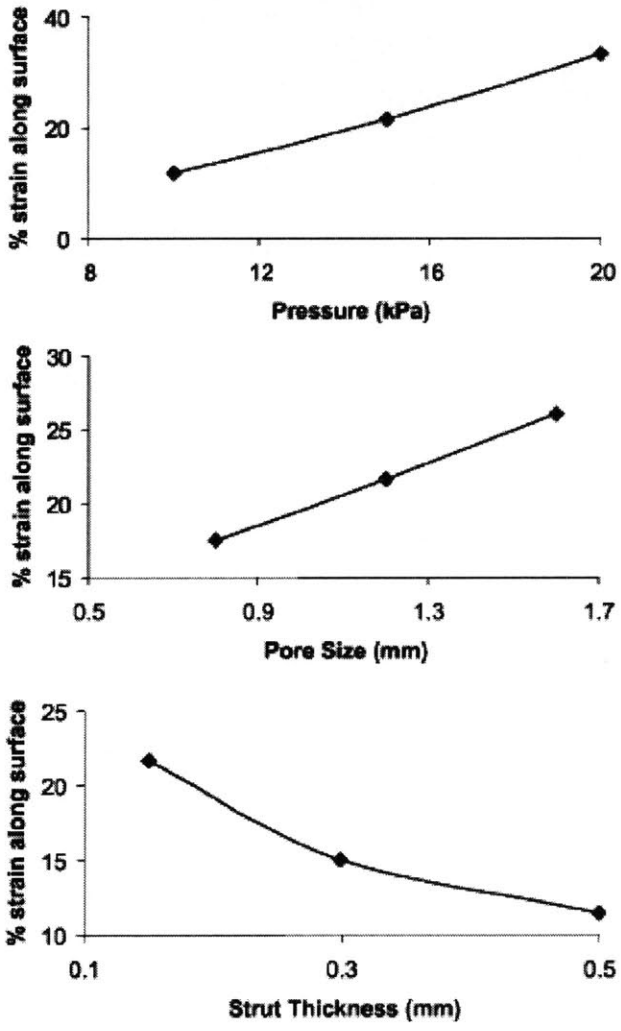


As tissue becomes fibrotic or edematous, the Poisson's ratio may increase, further decreasing the average and peak surface strains. In the simulation, raising Poisson's ratio from 0.36 to 0.50 (incompressible tissue) resulted in the decrease of average surface strains from 26 percent to 22 percent (Figure B above). In the ranges studied, surface strains also appear to be far more sensitive to changes in tissue stiffness than tissue compressibility. Our model shows that raising tissue stiffness by 40 percent causes a nearly 66 percent decrease in the surface strain, whereas an equivalent change in compressibility results in only a 15 percent decrease. This suggests that changes in the stiffness of the wound over time may be an important factor to consider when optimizing therapies for healing wounds.

Applied Strains Are Device-Dependent

It has been suggested that cells can be induced to respond to growth factors and proliferate when undergoing an optimal degree of strain [7], [51], [5], [52]. To examine how changes in VAC device properties can be harnessed to produce optimal wound tissue strain, finite element analysis models of the VAC were constructed with varying

pore diameter, strut thicknesses, and imposed pressures. Strain is very sensitive to changes in pressure. Doubling applied pressure from 10 kPa to 20 kPa also doubles average surface strain (Fig. C, above), but doubling average pore diameter from 0.8 mm to 1.6 mm results in only a 50 percent increase in surface strain (Figure C, center). In addition, decreasing the thickness of the sponge struts causes a decrease in the average surface strain (figure below—from [2]). Considered together, these results suggest an exquisite sensitivity of the imposed strain to VAC device properties and demonstrate a broad dynamic range of achievable strains in the wound tissue achieved by varying device parameters.



Conclusions from this study

We've already discussed in the introduction to the thesis the basis for our hypothesis that micromechanical forces are at play in the mechanism of action of the VAC device. We've shown that the strains obtained through the use of the finite element model do indeed approach physiologically meaningful values. Let's now discuss how the VAC device may be enhanced. We will then discuss some of the limitations of our model.

Enhancing the VAC Device

The finite element analysis results obtained here, particularly with respect to pore diameter and strut thickness, may be helpful in optimizing VAC device design. For instance, if higher strains are desired to facilitate cell proliferation, our data might suggest that both increasing the pore diameter and decreasing the strut thickness will increase overall average wound strains. Unfortunately, the model also implies that the higher average strains thus achieved will inevitably be associated with tremendous spikes in strains and stresses in regions immediatelyadjacent to the struts, which may cause local damage to the wound.

Moreover, larger pore diameters may also simultaneously result in larger areas with locally suboptimal strains. Indeed, our model suggests strains peak immediately adjacent to the strut, become optimal (~10 percent) farther away from the strut, and become suboptimal (<5 percent) with increasing distance. Optimizing sponge pore design thus becomes a delicate balance between minimizing the central suboptimal strain regions and minimizing high local strains caused by increasing the number of struts. Our finite element analysis results imply that some of these issues might be addressed through the use of stress concentration-reducing rounding of struts. A great deal of work in the biomechanics literature correlates changes in elastic modulus and Poisson's ratio with changes in tissue abnormality and injury, tissue type, and age-related modifications [49]. By combining these data with the kind of systematic analysis presented in this article, our model could be extended to different tissue types or even to the same tissue as it changes over time.

If the VAC device does work as we have postulated, then it may be possible to effect faster and more effective responses to wound healing by inputting forces directly to the wound bed. First let's try to understand how forces may be transmitted through the wound bed. The section 'A conceptualization of force-transmission to cells in the wound bed' in the first chapter shows a schematic which explains that it is not enough to simply apply forces from the outside on to the wound bed. Since the wound bed is essentially composed of weak connectivities, any forces that we apply will tend to be dislocated through only the stronger sections of the wound bed and the weaker sections will see essentially no forces or strains (the stresses will rise—since now we have the same forces acting on a smaller area—the larger global area because part of its links were severed—will reduce to a smaller effective area). Thus, we will have done essentially what we said above we should not do. We will have raised the principal stresses at the force carrying loci while reducing to zero forces at the severed loci.

This is in essence why we need to input *Micromechanical forces*, or forces that act at a local part of the wound bed at each part of the wound bed. By providing a scaffold that will essentially create a stronger bed over which the cells can then attach themselves to, we have created a stronger cross-sectional area, all of which can support the forces applied.

Some other factors that can help to design a better scaffold for the Vac Assisted Closure therapy include ascertaining whether constant strain is better (by fixing the applied strain—a fixed strain needs to be able to control the force at each micro-loci by inputting a fixed motion at each micro-loci regardless of resistance from the tissue), or is constant stress better (ie by applying a fixed force). For example, in a healing wound bed, the elastic properties of the tissue at each loci will not be the same. In other words, the tissue will display non-homogeneous behavior in terms of material properties.

Further when a force is input, the force has to lower than the yield stress of the weakest area of the wound bed. Otherwise the force will cause that area to expand until it breaks,

and that will stop any further force transmission through that axis. One idea is to use magnetic beads in a meshwork that can be moved through the application of an applied field. This method however may suffer from the same problems that we see with constant force application. If the density of beads is constant throughout the wound bed, then the beads will see the same magnetic force and therefore we again end up with a constant force boundary condition. However, if the density of the beads can be coupled to the material properties of the wound bed, then we can couple the force to the wound bed properties.

One way to do this may be to look for correlations between mechanical properties and the physical condition of the wound. For example, in regions where the wound bed is weaker, we may see more fluid material. Magnetic beads that don't disperse as well in a more fluidic phase may provide a solution (although this is just an example—I doubt if such particular beads exist).

Another (simpler) technique is to use a material that expands with temperature or with some other input (such as a voltage—a piezo-electric material for example).

In both of the above uses, we would need to be able to separate the material from the species (such as cells) in the wound bed that can couple to the material. This can be accomplished by using connecting ligands on the material that can bind cells, but that can be cleaved from the material (but not from the cells). Thus, once the piezo electric material has fulfilled its task of force application, it can then be removed from the wound bed.

It would seem reasonable to assume that if the forces are applied uni-directionally with time, then the collagen fibers and other structural components will align along the applied forces. This usually is the reason for the strong retractive forces seen in scar tissue.

Anisotropy (ie directional nature of the mechanical properties of a material) is thought to confer structural strength in the case of myocardial tissue. Further "protein turnover and

myofibrillar structure are regulated differently by stretches applied parallel or perpendicular to the long asis of aligned cardial myocytes"..."suggest[ing] that both cellular prealignment and biaxial stretch protocols will be required to optimize directed growth and remodeling of evolving engineered heart tissues [53].

Limitations of the Model

Although biologic tissue is a nonlinear viscoelastic system,15 our model was constructed as a linear elastic system. The linear assumption was deemed reasonable, as the tissue stress-strain relationship remained linear over the relatively small range of strains on most elements. Destruction of tissue architecture was assumed to render the normally *anisotropic* material properties of healthy tissue close to isotropic in the wounded tissue. Furthermore, the biologic system is dynamic, and local cell growth partially relieves stresses that occur in wounds. These aspects of the wound-healing process were not studied in our model. We assumed the struts in this model to be fixed in space and completely incompressible. However, as the actual struts likely have a degree of compressibility, the model may overestimate the peak strain spikes near the struts of the VAC device.

Extension of this work to gene-expression study

Now that we've seen that the strains obtained in a wounded model of the skin are significant, we next study the gene-expression profile changes that may occur in a study that studies the effects of forces on normal tissue alone. A future extension of this work will look at the effects of forces (in the context of the VAC device for example) on wounded tissue. Such samples are in the process of being generated.

The next chapter thus is a study of the gene-expression profile changes seen with the application of forces to rat ears.

Chapter 4

Effects of forces (stretch forces) on the gene expression profiles of rat ears

With the analysis in the previous chapter showing us that significant forces are imparted to a tissue subjected to the VAC device and with the additional evidence that the VAC device does show substantial improvements in wound outcome, we next study at the gene expression level what types of genes (and therefore by extension their products, proteins) are differentially expressed in a tissue subjected to a force.

As explained in chapter two, we decided to use the rat ear in our exploration of the effects of forces on normal tissue. An extension of this thesis will look at the effects of forces on wounded tissue (rat ear in our experiment, even though it is stretched with non-physiologic forces is assumed not to be wounded because we have not induced any external tear on the ear).

Prior work done at Dr. Orgill's lab by others has shown that rat ears exposed to forces show a growth in the number and size of the vasculature draining the ear.

With the above work accomplished, we would now like to explore this mechanism at the genetic level by decoupling it from the way the VAC obtains the strains. In other words, we would now like to apply mechanical forces directly to tissue to see what effect this has on the tissue. *A gene expression analysis on a suitably excised ear tissue from each model is then be conducted to look for patterns of gene expression.*

Work done by Dr. Orgill and co-workers has shown that a rat ear exposed to a constant strain becomes more angiogenic—with the vasculature showing significant growth both in size and density following the application of a strain or force. An intermittent application of force has been shown to show even better outcomes with the VAC device.

It is also known that cyclic tension increases the rate of breast tissue growth. Thus an extension of the constant force work is now being studied by Dr. Orgill and Co-workers to see whether a varying force gives superior or different results. *To start with, however, in our gene expression study, we will study only the effects of constant forces.*

A validation of this work can be carried out by essentially 'knocking out' the gene or geneset (we'll discuss genesets further down) that seems to show promise through our analyses.

Importance of studying gene expression

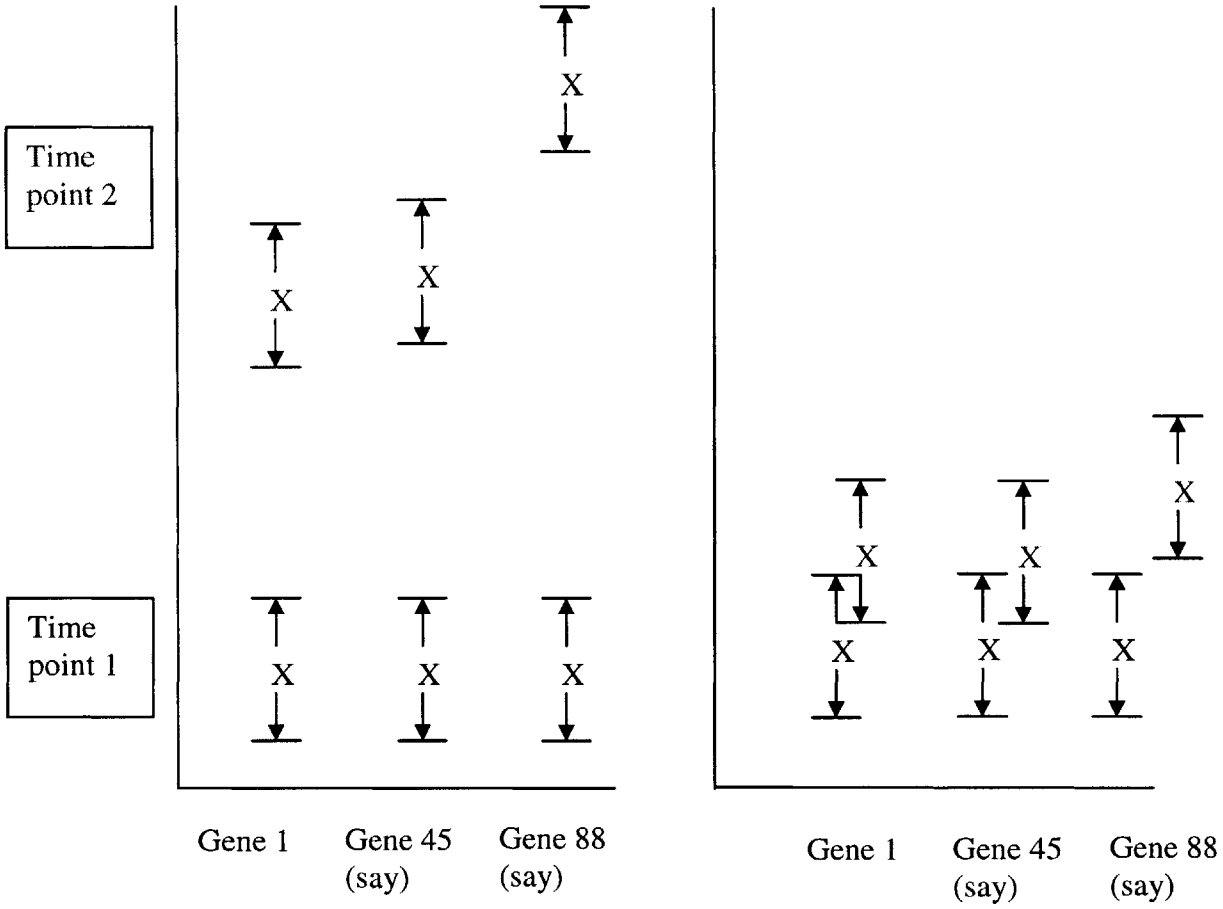
The goal of gene expression studies is to find a set of genes that show differential expression with respect to the experimental conditions. Genes that are differentially expressed and whose expression level changes are correlated are believed to be part of the same biological pathway. Thus gene expression profiling can give us an understanding of the pathways underlying the process being studied.

Analyses (typically) used to study gene expression

One popular technique for obtaining important genes relevant to a system is hierarchical clustering. This technique is especially powerful when applied to a system that is clean and simple (for example free of too much noise and where expression level changes are not so subtle). One metric used in hierarchical clustering is the Pearson Correlation coefficient. However, when the system being studied is 'messy' then hierarchical clustering can give us results that may be misleading.

If for example, the results of an expression study give us a high fold changes in a set of genes, then these genes will quickly stand out from the noise (figure on left) and expression changes will speak about similarity between genes (noise is given by the

arrows). We will then conclude (for example) that genes 1 and 45 are closer in expression changes.



Other analyses that are used to obtain important genes and gene clusters (or sets) are self-organizing maps, principal components analysis etc.

These types of analyses are not useful in our dataset

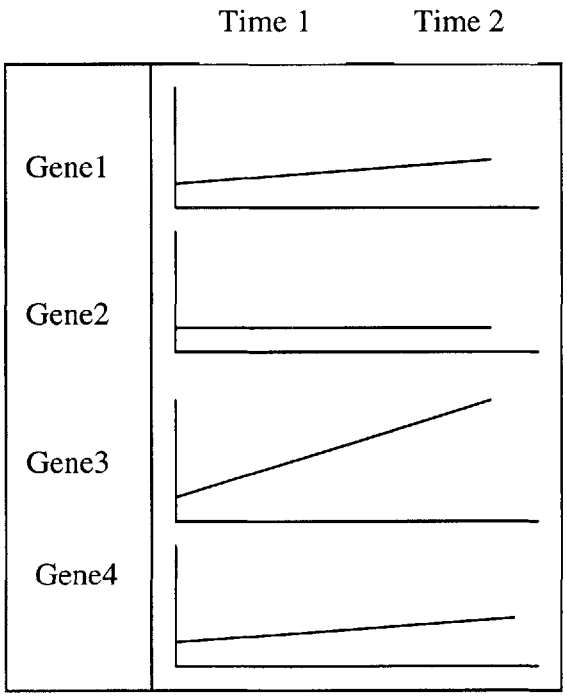
If, however, as we see in the figure on the right, the expression changes are subtle, then depending on where in the value spectrum given by the noise the data point lies, we may end up saying that gene 1 correlates more with gene 88 (say) and not with the gene that it

is really correlated with (gene 45) if there is no noise. Thus, we need a different way of extracting important genes and gene clusters.

Because many gene sets and pathways have over time been extracted from many simple and clean experimental systems, we can test different gene sets and see which ones seem to be the most strongly 'enriched' in our system. This is where GSEA or gene set enrichment analysis enters the picture.

Hierarchical clustering cannot be used in ranking the genes in GSEA

We rank the genes in terms of two data points (or sets) by taking the difference between the two points in expression values. This gives us a range of absolute differences. However, hierarchical clustering will only look for closeness in expression between two points. Thus two points which are closer in expression space will be placed 'highest'. As we've discussed above, if the noise masks subtle expression differences, the wrong genes will end up being correlated. Thus, a simple difference in expression levels (or in our case, the 'difference paired t statistic squared') will give us a ranking that is not 'influenced' by noise effects.



Now if we rank the genes according to difference, we'll get Gene3 > Gene4 > Gene1 > Gene0. However, if we cluster hierarchically by correlation, genes 4 and 1 will be placed together at the top since their expression levels *patterns* are the most closely aligned.

Time series analysis of rat ear stretch

To keep a continuous thread to our story, I will begin at the beginning, where I applied stretch to rat ears and then expression profiled the ears. It may help to keep in mind that we are trying to understand the effects of forces applied to rat ears on the gene expression changes that occur in the rat ears. Here's an overall summary of the scheme:

1) Apply stretch to rat ears
2) Excise the ears and immediately freeze them in liquid nitrogen at different time points
3) Store the frozen ears at -80C
4) Extract RNA from the ears
5) Hybridize this RNA to affymetrix RAT chips 230 2.0
6) Obtain expression data
7) Conduct GSEA on the expression data
8) Find groups of highest enriched genesets (or rather probesets) that also have a low P value (computed with the permutation testing)
9) Check which genes correspond to the probesets in each geneset
10) Look for a biological mechanism that may help to lend support to the enriched geneset

A final follow up of the above may be to conduct further biological validation steps to test the biological significance (if any) of the results.

The rat ear stretching experiment

The rat ear was chosen to study the effects of forces on tissues because it is relatively thin (and therefore easily visualized under the microscope). It is also easy to apply forces to the ear. Further most of the ear is skin (with a thin internal lining of cartilage). The rat ear is also fairly large (compared say to a mouse ear) and therefore it is easier to handle (pigs are bigger but they are far more expensive and cumbersome to maintain).

Also, previous work at the lab had been conducted on the rat ear to study the effects of forces on the size and number of blood vessels in the ear. Thus, there was a pre-existing model that could be further modified to perform our study.

The ear gluing and stretching system

For this experiment, it was important to have a system that could apply a stretch force on the rat ear without explicitly wounding the ear. This is because, as mentioned earlier, we are trying to decouple the gene expression profile under forces from the gene expression profile under forces on a wounded tissue. This way, we obtain an understanding of the effects of forces independently of tissue wound.

Thus, we could not suture the ear to any force producing mechanism, because the suture would create an inflammatory response that could potentially overwhelm any other response. A mechanism that had been worked out in Dr. Orgill's lab used latex taken from a lab glove to glue the rat ear to an externally tethered mount. Superglue was used in that technique to bind the ear to the latex piece. However, we needed to keep the ears glued for an extended period of time and so we had to search for other methods that could give us longer lasting binding. Superglue is ethyl cyanoacrylate. However dermbond is Octyl-cyanoacrylate and its application gave us far longer gluing properties. The stretching system is described next.

One solution to the ear gluing problem

When items are glued together, they have to resist two types of forces: shear forces and perpendicular forces (forces that tend to pull the two pieces apart in tension). By using a surface with a high friction coefficient, we can add to shear resistance of the glue, since frictional forces will aid in the strength of the glue in shear. It needs to be remembered that glue acts through bonding interactions and not just Wan Der Waals/close electrostatic interactions while friction acts through the latter. One idea to increase a glue's holding strength is to use a clamp around the piece (say ear) to be glued to with a material (say two pieces of latex). If the clamp is rigid, then the normal forces acting on the latex pulling it off the ear go almost to zero, thus leaving shear forces as the sole forces holding the ear to the latex.

Further, if the clamp is slightly pushed inward, frictional forces are increased highly since friction depends on the normal force with which the surfaces contact.

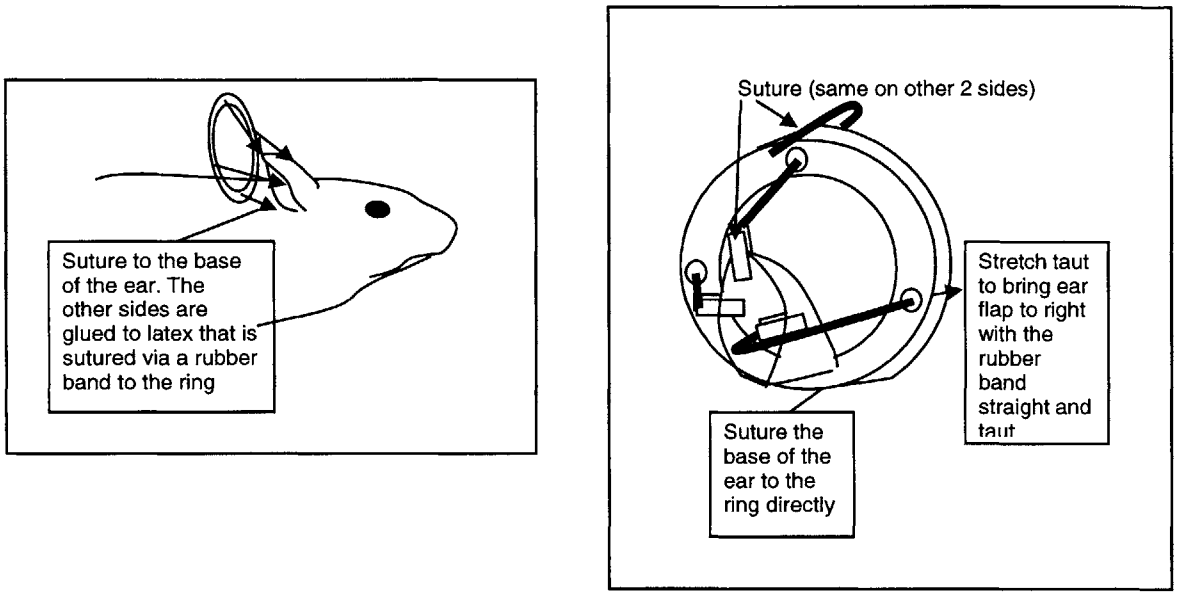
This system however was not used in our study. For one, we were able to obtain dermabond which by itself held the rat ear in place with the latex for long enough periods of time. Further, some of the rat ears showed tearing. This could have been because the clamping forces applied were too strong. This technique however held even with the weaker superglue (without use of the dermabond) even when the ears were partially or completely torn off.

The stretching experiment

Male Wistar Rats (scientific name Norvegicus Ratticus) were used in all our analyses (mixing males and females would have entailed possible impracticalities with pregnancies). The rats were ordered at 150g each. We decided to study the effects of forces at 10 discrete time points. Because earlier time points were probably more important in the response to mechanical stress, 7 of our time points were taken in the first 24 hours post stretch. Three rats were used for each time point. One ear of each rat

(chosen randomly) was stretched and the other ear was left unstretched (this became our control).

The overall stretch apparatus is shown below and the scheme for obtaining stretch is shown two figures down.



Plastic caps from were obtained. The rat ear is shown in better detail on the right side of the figure. The rats were allowed to acclimatize at least overnight to their new surroundings.

In preparation for the stretch apparatus application, the rats were anesthetized with ketamine and xylazine (for an approximately 200g rat, use 0.2ml Ketamine and 0.15ml Xylazine). The anesthesia was injected into the lower abdomen. The angle of the syringe was kept at a low angle so that the anesthesia was placed into the peritoneum (but not so small as to be constrained in the overlying musculature—anesthesia flows faster out of the peritoneum than it will from overlying muscles) in the lower abdomen right above the bladder region of the rats.

During the procedure in which the apparatus was applied, the rats were given maintenance doses of ketamine (maintenance doses of ketamine—0.06ml—are given every 2 hours or as needed if the rats seem to be waking up).

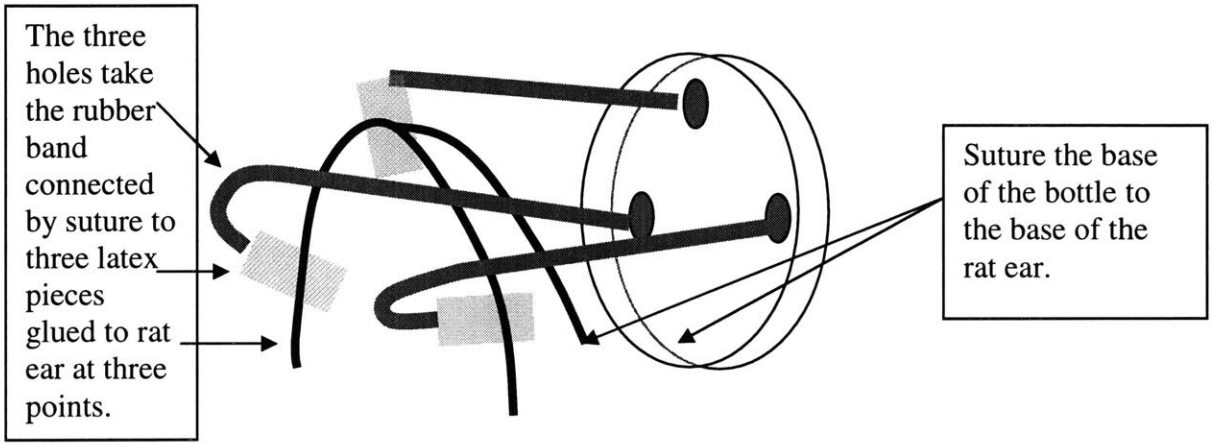
Once the rats were anesthetized, their lower backs were shaved with a mechanical hair trimmer. At this location, a number identifying the rat was written with a marker pen. Then, a depilatory was applied to the ears and left on for 4 minutes after which it was wiped off first with dry paper napkins and then with ethanol.

Separately Latex gloves (specific company) were washed with soap (to remove any powder that may have been present on them) and then cut in strips that were 5mm wide and 14mm long (check on this). The strips were left to dry. Usually latex gloves are coarser on the outside (the side used to grip objects). Glue was used on this side to attach the strip to the ear.

Dermabond (Octyl cyanoacrylate)—which sets very fast (in a matter of minutes) was used to glue the strips to the ear. Each strip was applied at three points on the rat ear as far spaced as possible. Each strip was folded such that half bonded to the front side of the ear and the other half when folded bonded to the back side of the ear (as shown in the figure above—it almost looks as if there are two strips per point but that is shown only for clarity purposes.

A little bit of space was left between the folded parts of the strips (in other words out of the 14mm, only about 5 or 6 mm actually contacted each side of the ear). About 2mm was left away from the ear. To this 'jutting' part, we connected a suitably calibrated rubber band using sutures. The rubber band was calibrated using a tensiometer such that 50g of force was applied to the ears.

A plastic bottle cover was obtained (put down name of the bottle brand here). Three holes were drilled into the cover as shown below.



Once we've attached the latex to the rubber band through sutures, we pass the rubber band through its corresponding hole in the plastic cover and then measure the amount of stretch the rubber band obtains equal to that needed to give us a force of 50g (the rubber band is calibrated by marking 1mm marks on its unstretched length and then stretching it until the tensiometer shows 50g—this stretched length is then measured—this is the length that the rubber band is then later stretched to give us the requisite 50g force).

After measuring the stretch in the rubber band we then clamp the rubber band at that stretch by using a titanium clamp (we also devised our own clamp by using staples—we extracted free staples and cut them in half, and then used each staple within a titanium clamp 'applier' to clamp the rubber band—the clamp 'applier' is essentially a scissor like implement that can 'squeeze' the half staple onto an object as shown conceptually below).

Three rats were used per time point (giving us three biological replicates). One ear was stretched, the other ear had exactly the same mechanism applied except that stretch was not applied. This ear (chosen randomly) formed the control ear.

Time point	1H	2H	4H	9H	14H	19H	24H	48H	72H	96H
Stretch	3 rats one set of ears	3 rats one set of ears	3 rats one set of ears	3 rats one set of ears	3 rats one set of ears	3 rats one set of ears	3 rats one set of ears	3 rats one set of ears	3 rats one set of ears	3 rats one set of ears
Control	Other ears from 1H rats	Other ears from 2H rats	Other ears from 4H rats	Other ears from 9H rats	Other ears from 14H rats	Other ears from 19H rats	Other ears from 24H rats	Other ears from 48H rats	Other ears from 78H rats	Other ears from 96H rats

Once the stretch mechanism had been applied, an elizebethan collar was placed around the neck of the rats to prevent them from accessing the mechanism with their forelimbs. We kept only one rat per cage as we found that if we kept two to a cage, the rats 'helped' each other remove the collars!

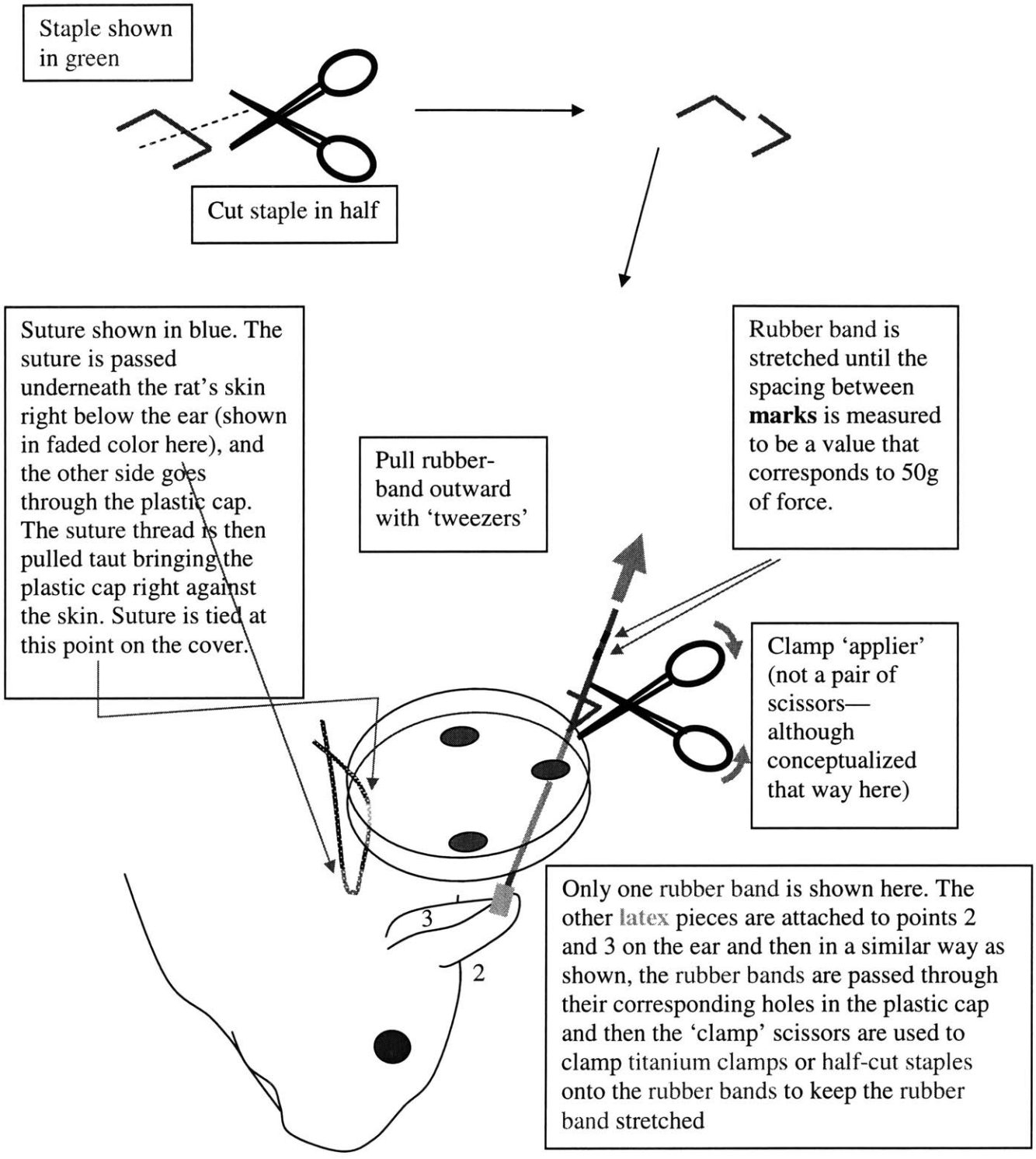
After the collars had been placed around the rat necks, we sutured the plastic covers from each ear to the collar gently. This kept the ears spread outward slightly and prevented the rats from dragging their ears on the material in their cages.

The rats were given buprenorphine (0.03mg/ml) to relieve their pain from the stretch.

The rats were then returned to their cages. Buprenorphine was administered twice daily for 4hours (for those rats that were used for longer time points) after which it was stopped.

Approximately 30 minutes before the time to sample the rat ears (which essentially meant that the rat ears were excised with a surgical scalpel), the rats were anesthetized (it was necessary not to euthanize the rats before excision as this could affect the gene expression profiles that would be obtained from their ear tissue.

Liquid nitrogen was obtained and using a pair of tweezers, the excised rat ear was directly inserted into the liquid (it is important not to bring any part of the body in direct contact with the liquid as this will damage the area—especially if the contact is prolonged). The frozen ear was then stored in an eppendorf that was suitably marked and the eppendorfs were then stored at -80C until they were ready to be taken to the RNA extraction and hybridization facility. All samples were kept in dry ice after they were frozen and when they were not in the -80C freezer.


Staple shown in green
Cut staple in half
Suture shown in blue. The suture is passed underneath the rat's skin right below the ear (shown in faded color here), and the other side goes through the plastic cap. The suture thread is then pulled taut bringing the plastic cap right against the skin. Suture is tied at this point on the cover.
Pull rubber-band outward with 'tweezers'
Rubber band is stretched until the spacing between **marks** is measured to be a value that corresponds to 50g of force.
Clamp 'applier' (not a pair of scissors—although conceptualized that way here)
3
2
Only one rubber band is shown here. The other latex pieces are attached to points 2 and 3 on the ear and then in a similar way as shown, the rubber bands are passed through their corresponding holes in the plastic cap and then the 'clamp' scissors are used to clamp titanium clamps or half-cut staples onto the rubber bands to keep the rubber band stretched

BECKMAN DU-600 Date: 08/16/04
Time: 02:23

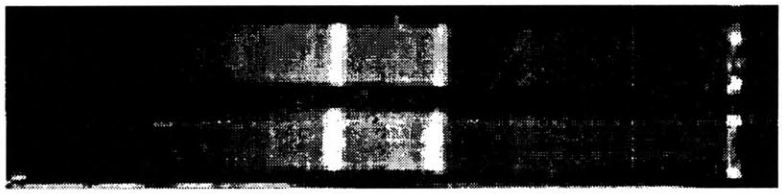
DNA/Oligo Quant
ReadSamples RawData Method SaveClear Print Quit

Results file: A:\WORK_RES Method name: A:\RNA

Assay type: RNA Background corr: [No] 320.0nm
Sampling device: None Pathlength: 1.0000 cm
Read average time: 1.00 sec Conc. factor: 40.000 at 260.0nm

Sample ID	Net Abs 260.0nm	Net Abs 280.0nm	260.0/ 280.0	Dil. Fact.	Conc. ug/mL
1	0.0004	-0.0001	-3.6365	100.00	1.7830
C	0.6816	0.3198	2.1313	100.00	2726.5894
3	0.6845	0.3208	2.1334	100.00	2738.0432
S	0.6039	0.2836	2.1294	100.00	2415.5818
5	0.6051	0.2841	2.1300	100.00	2420.3608
6				100.00	

RNA integrity on gel



Gene set enrichment analysis

In their paper [54], Mootha et al use GSEA to find gene sets that are important in diabetics. They didn't perform a time series analysis. If they had performed a time series analysis, they would have had (say) a set of time points of diabetics over normal controls. Thus, our data is really analyzable in two days. We can perform an analysis that performs GSEA analysis over all time points comparing stretch with controls. This analysis would

test whether given that a gene set may be important in our system, whether there is in fact a difference between control and stretch conditions.

If, however, we normalize the data between stretch and control by taking their ratios, an SVD then on the data matrix may give us insights into the genes that may be important over the time course of the experiment.

We can think of two types of genes (or gene sets—we'll define this later) that may be affected (albeit they could turn out to be the same) between control and stretch conditions and those that may be affected over time. So let's say that the effects of stretch are time independent. For example, if gene A is upregulated between control and stretch but stays at the same level of upregulation over time, then if we normalize between control and stretch for each time point, the ratio of stretch to control for gene A will stay constant, and gene A will not pop out of any analysis of variance. However, if the ratio shows a shift over time, then an analysis of variance will pop out this gene.

Mootha as we explained above had eight replicates at the same time point. Thus they could average their control and stretch conditions and take the differences to rank their genes. However, we cannot do this. We have a time series and each pair of control and stretch expressions. First, let's get an overview again (steps 7 and on in the previously explained scheme):

1. Obtain the Rat ear expression data set
2. Compile gene sets (we'll explain in detail further down how we did this)
3. Conduct GSEA
4. Perform Permutation testing

The first step was performed at the Children's Hospital RNA expression facility. Some of the key steps are explained in the Appendix.

A gene set is a collection of genes that are somehow related. They can for example be part of a common pathway or they can be part of the same putative pathway.

A number of the gene sets that were used were obtained from Mootha's previously compiled and reported genesets reported in [54] and located at [55]. However, the probesets compiled by Mootha are human HU-133A (a type of human genechip) probesets. Thus, we had to convert these probesets to rat orthologs. We can convert Affymetrix probesets from one genechip type to their orthologs at the Netaffx analysis center at the Affymetrix website (simply by doing a batch query and outputting orthologs).

Other genesets were obtained from geneontology [56] by typing key terms that may stand for important pathways (eg. Angiogenesis, hypoxia etc). GeneOntology then output a set of different 'hits' with relevantly named 'gene lists' or 'gene sets'. These links then gave us the genes under that pathway.

However, our expression data is represented by probeset ID's not genes. A probeset is a term for a set of 'probes' that represent a gene. They are 15 to 20 nucleotide long probes that interrogate pieces of RNA that are obtained from the sample. The probes are chosen so that they can uniquely identify one single gene [57]. Each Affymetrix genechip has probesets for the genes that that chip can interrogate. There are multiple chips even for each species.

As was mentioned above, Mootha's data are replicates at the same time point. Thus, there are no pairwise comparisons to be made. In our case, we needed to conduct a pairwise test. Further, we couldn't just take each time point (control and stretch) and obtain anything meaningful from it, because after we run our GSEA analysis, we need to be able to run a P value test that will give us an idea of how significant our gene set analysis is. To run the P value test, we need to be able to shuffle our columns. We have eight control and eight stretch columns. If we shuffle randomly, we'll on average get C(8,4) number of shuffles or combinations which is [8*7*6*5/(4*3*2)]^2 = 4900 permutations which is

well above 1000 permutes. If however we only had 7 time points we'd be short of being able to obtain 1000 permutes. We'll explain further down how the permutations were in fact carried out.

Therefore we decided to use the paired difference T statistic to rank our gene list. The paired T statistic is given by,

$$t = (\overline{X} - \overline{Y}) \sqrt{\frac{n(n-1)}{\sum_{i=1}^{n} (\hat{X}_i - \hat{Y}_i)^2}}$$

To make computation easier, we squared this statistic and then ranked the genes. In the above equation, $\overline{X}$ is the average value of the control set across a gene, and $\overline{Y}$ is the average value of the stretch set across a gene. $\hat{X}_i = X_i - \overline{X}$ and $\hat{Y}_i = Y_i - \overline{Y}$. 'n' is the total number of entries of control or stretch for each gene (8 in our case).

We can perform two fundamentally different types of analyses if we keep the sign of $\overline{X} - \overline{Y}$ versus if we don't. Squaring the t statistic removes the sign, but we can always insert it back in. Let's see what keeping the sign means.

Using the t statistic squared with the sign

If the stretch condition overall has a higher expression value versus the control set, then $\overline{X} - \overline{Y}$ should give us a negative sign. Similarly, if the stretch condition expression is lower than control, then $\overline{X} - \overline{Y}$ is positive. Thus, if we're looking for doing an analysis where we try to rank genes by their degree of upregulation in the stretch condition, then we should keep the sign in our squared t statistic, and we should rank the genes such that the most negative values of the t statistic squared (with the sign kept) go to the top. While if we want to do a downregulated analysis, we should reorder our genes such that the most positive t statistic squared value (again with the sign of $\overline{X} - \overline{Y}$ kept) going to the top.

Mootha's analysis [54]in fact only ran the first upregulated case mentioned above. However, there can certainly be value in running the analogous downregulated analysis as it can give us important insights into what genes (or gene sets and therefore pathways) are suppressed in the system under study.

However, we can go one step further. We can in fact do a combined analysis where we look for changes in expression regardless of sign. If the changes are big, then those genesets should go to the top regardless of whether the genes in the genesets are up or down regulated. Let's study this way of analyzing the data now

Combined up and down regulated gene set enrichment analysis using the paired difference t statistic

By not keeping the sign in our t statistic squared, we're now essentially looking for aggregate largest changes regardless of whether the changes come from up or down regulation. For example, we may have a geneset that has members some of whom are highly upregulated while some others are highly downregulated. Such a geneset may not come up to the top in a simple up regulated or a simple down regulated analysis. However, it may come up to the top if we do a combined analysis. For example (we'll go into detail about this later), we found that hypoxia came up to the top using such a technique. Using either the up regulated or the down regulated analyses alone would not have allowed us to find this gene set. It may be highly important because its members are *most* differentially expressed. However, their different signs mean that they don't show up at the top of the signed analyses.

Once we rank our genes using either of the above analyses, we need to find a score for the different gene sets that we have that will give us an idea of the importance of that gene set (or pathway) in our analysis. Let's take a look at how such a score is found.

Finding the enrichment score (this is a normalized Kolmogorov-Smirnov Statistic [54])

Let's say next for illustration purposes that our data matrix has genes labeled 1 through 10 (our real dataset has 31099 genes or probesets). Let's say this geneset has within it genes labeled 2, 7, 9. The unordered genes look like the following with the second column showing hypothetical values of the t squared statistic calculated.

Gene label	t statistic squared
1	0.4
2	0.8
3	0.2
4	0.7
5	0.9
6	0.3
7	0.6
8	0.4
9	0.1
10	0.6

Let's say we want to find the enrichment score or ES for our geneset. Let's say that our genes have been reordered based on the t statistic squared column. The following table shows this graphically. The first two columns are reordered using the second columns. The rest of the columns show conceptually how the enrichment score is obtained. In the third column, we've marked where the gene set members are located. For clarity, the fourth column has shown the formula for the Kolmogorov statistic. Note that the statistic value is the same number for genes that are not members of the gene set group, and another same number for members of the group. The statistic is then obtained with G being 3 (number of genes in gene set) and with N being 10 (total number of genes in our analysis).

The sixth column then computes a running sum starting at the first row. The enrichment score is the maximum value obtained by this running sum.

Gene label	t statistic squared	Gene set	Kolmogorov Statistic	Statistic value	Running sum	ES
Gene 5	0.9		$XN = -\sqrt{\frac{G}{N-G}}$	$XN = -\sqrt{\frac{3}{10-3}} = -0.65$	-0.65	
Gene 2	0.8	x	$X = \sqrt{\frac{N-G}{G}}$	$X = \sqrt{\frac{10-3}{3}} = 1.53$	0.87	
Gene 4	0.7		$XN = -\sqrt{\frac{G}{N-G}}$	$XN = -\sqrt{\frac{3}{10-3}} = -0.65$	0.22	
Gene 7	0.6	x	$X = \sqrt{\frac{N-G}{G}}$	$X = \sqrt{\frac{10-3}{3}} = 1.53$	**1.75**	←
Gene 10	0.6		$XN = -\sqrt{\frac{G}{N-G}}$	$XN = -\sqrt{\frac{3}{10-3}} = -0.65$	1.09	
Gene 1	0.4		$XN = -\sqrt{\frac{G}{N-G}}$	$XN = -\sqrt{\frac{3}{10-3}} = -0.65$	0.44	
Gene 8	0.4		$XN = -\sqrt{\frac{G}{N-G}}$	$XN = -\sqrt{\frac{3}{10-3}} = -0.65$	-0.22	
Gene 6	0.3		$XN = -\sqrt{\frac{G}{N-G}}$	$XN = -\sqrt{\frac{3}{10-3}} = -0.65$	-0.87	
Gene 3	0.2		$XN = -\sqrt{\frac{G}{N-G}}$	$XN = -\sqrt{\frac{3}{10-3}} = -0.65$	-1.53	
Gene 9	0.1	x	$X = \sqrt{\frac{N-G}{G}}$	$X = \sqrt{\frac{10-3}{3}} = 1.53$	0 (round off neglect)	

We see in the above that the ES is 1.745743 (the highest running sum number). We see that if all the members of the gene set are at the top of the list, then our ES will be a maximum value. Second note that the Kolmogorov statistic can really only take two values: one value when the gene is part of the gene set defined by 'X', the other value when the gene is not part of the gene set XN (note also that this value is negative). Further, X and XN are reciprocals of each other with XN also being negative. The running sum always ends up at zero. We can see that if the gene set members were all at the top with no non group member in between them, then the ES would be max at the last member of the group. If all the gene set members were at the bottom, then the max ES would be zero.

Now, we'll take another example to find the ES for another geneset which has 2 values – genes 5 and 4(remember the number of genes in a gene set can never be greater than the total number of genes in the analysis. Also note that all genes in the gene set have to be found in the total number of genes that we have. A good way to have a check on this is to make sure that the running sum finally is zero at the last row (the reordered table above has gene # 9 at the last row and we see that the running sum is zero).

Our new values for XN is $-\sqrt{\frac{G}{N-G}} = -\sqrt{\frac{7}{10-7}} = -1.527525$

and for X is $\sqrt{\frac{10-7}{7}} = 0.6546537$. We'll again do the running sum and the ES for illustration purposes. We already have our (hypothetical t squared statistic calculated—this is only computed once). The running sum and ES for gene set 1 (we're calling the gene set we started with gene set 1 now) are copied from the table above. Gene set 2 is shown next. We've deliberately chosen the gene numbers in gene set 2 to show that the higher up the gene set members are, the higher the ES will be, and this is independent of the gene set size (gene set two is smaller than gene set 1). The calculations for gene set 2 are shown in grey.

Gene label	t statistic squared	Gene set 1	Running sum	ES	Gene set 2	Kolmogorov statistic value	Running Sum	ES
Gene5	0.9		-0.65465		x	2	2	
Gene2	0.8	x	0.872871			-0.5	1.5	
Gene4	0.7		0.218218		x	2	**3.5**	←
Gene7	0.6	x	**1.745743**	←		-0.5	3	
Gene10	0.6		1.091089			-0.5	2.5	
Gene1	0.4		0.436435			-0.5	2	
Gene8	0.4		-0.21822			-0.5	1.5	
Gene6	0.3		-0.87287			-0.5	1	
Gene3	0.2		-1.52753			-0.5	0.5	
Gene9	0.1	x	0 (round off neglect)			-0.5	0	

The above (hypothetical) analysis can be extended to any number of gene sets. Looking at gene sets rather than individual genes reduces the noise in our system by a factor of $\sqrt{G}$ where G is as above the number of genes in a gene set [54].

In our system, we had 173 gene sets. We'll explain the findings in our results section in this chapter.

After computing the ES for all the gene sets in our system, we look for the gene set with the highest ES. This gene set (or pathway) then is our most important pathway in our analysis. However, this is not the end of the story. Usually when we find a certain finding, we want to be able to state what level of confidence we have in our finding, that the finding is in fact real and not the result of a chance occurrence. In other words, we want to find the P value of our analysis. We want to say what is the chance that the gene set we obtained (the one with the highest ES) is a chance hit.

Carrying out the permutations

The name, permutation test, can be misleading. It may seem to imply a 'simple' permuting or shuffling of the data. That is not the purpose of the permutation test. In a permutation test, the goal is to be able to discriminate between two conditions. Are the differences between two conditions real or are they the effect of chance occurrence? Therefore the 'shuffling' that needs to be performed has to be able to 'ablate' the structure in the data that gives us a difference only if the difference is real. If the difference isn't real, the ablation will do nothing or will do far less.

Let's see how this can be applied to our system.

Our system is composed of two rectangular matrices (one if we normalize the data)—one is the control matrix, the other is the stretch matrix. We want to be able to tell if the stretch matrix condition in a certain row (say corresponding to gene 'x') is really different from the control matrix condition in the corresponding row (also pertaining to gene 'x'). Now if there really is a difference between the two groups, then in essence this difference will be a greater difference than we would have if we shuffled a few members between the groups, because by doing so, we're averaging out the 'real' difference that we see (this idea follows along the lines presented by Good [58]).

If however, there is no difference between the groups, any shuffling will not really produce much difference. In fact, the shuffling may actually produce more of a difference if the right numbers come together. Thus, the 'random' shuffling of the class members (or labels as some call them) may end up producing 'real' differences. This is the idea behind the permutation test. If the random shuffling of class members is able to give us significant differences enough times, then that means that the difference that we observe in the original data aren't real—they can be the product of a chance occurrence. Further, this type of a test is able to discriminate between real differences (by quashing the difference in a random shuffle) versus artifactual (or putative differences) that aren't quashable (and in fact may be exacerbated) in the shuffled data.

The above reasoning also answers why 'rearranging' the rows in our data doesn't do what we want it to do. Rearranging the rows cannot discriminate between real differences and supposed differences. It is not testing for differences between GROUPS. Remember that we're ranking the genes by some metric. The ranking is the RESULT of the operation of this metric. The metric is supposed to be able to identify differences between the groups (control versus stretch).

The value of the metric then gives the ranking order. If we were to simply shuffle the ranking order, we've really not asked to test for a difference between the two groups. This rearrangement will always make the data look like it is randomized. In that sense (see later) all gene sets will find themselves at the top some portion of the time. Rather than testing for a chance occurrence of a geneset coming to the top because it really isn't able to discriminate between groups, the chance of any geneset coming to the top becomes equal to any other!
In the first column we have the genes (one through 31099). The eight columns after the first are the eight time points we have.

Why do we keep each condition together? In other words, why do we swap whole columns rather than one member of a column only? Because in one sense each column is

a 'condition'. When we test a drug for its effects on sick people versus the effect of a placebo on another set of sick people, our subjects are people. We can think of this person as being comprised of many many genes (which he is—but we're thinking of the genes in a way where we think of them as forming a complete group). For example, if we wanted to test the strength of a table in terms of its being able to withstand a certain weight, we consider each part of the table as moving together. For example, let's say we have 6 tables (3 in one group and 3 in another), and we identify 12 'genes' or 'parts' of each table as 'identifiers' that can possibly speak to the strength of the table, when we 'shuffle' the tables between the groups to see if they are really different, we are really not allowed to shuffle parts of one table with parts of another table. This is because the property that we're looking at (strength) is a reflection of how each part of the table is connected to the rest of the table.

Analogously when we say we have six people, 3 women and 3 men, and we're looking to see if there is really a strength difference between the two groups, we're not really allowed to give one man's legs to a woman, because then we really haven't tested for differences between the WHOLE man and the WHOLE woman.

According to Phillip Good [58] in his book on permutation tests, first we identify the hypothesis and the alternative of interest. In our case, the hypothesis could be that stretching the rat ears really doesn't do anything versus the control case. The alternative hypothesis is that stretching does do something.

Next, according to (1), we need to choose a test statistic. We chose the t statistic. We calculate the test statistic AND then rank our genes. This ranking gives us our VALUE that we will study further (the value really is a record of all the genes in their respective positions with the rank reordering).

What essentially Good is saying is that if there really is a difference between the two conditions, then this VALUE will be an extreme condition, and a mixing of the samples will never replicate (or very rarely do so) the extreme condition. However, if the

difference is really not there, then a mixing will be the SAME as if we never mixed the labels (the label is really the one test condition) and we will replicate our reordering many times possibly.

Possible problems with the way the enrichment score is computed

But here's the real problem with the way we're doing the permutations. We're looking for the number of times a certain geneset goes to the top. We're not looking for how far away from the extreme its enrichment score changes. For example, let's say geneset A gets an enrichment score of 100, and let's say that swapping labels shifts the difference between the two conditions somewhat. However, if the enrichment score of geneset A is much above the next geneset (say geneset B), its 'diluted' ES score may still turn out to be far higher than the 'possibly increased' enrichment score from geneset B. Thus, what we should really look for is a shift away from the enrichment score for each geneset in each of the 1000 permutes rather than the maximum enrichment score obtained by each geneset in each of the 1000 permutes.

If two things are really not different, then switching things will not change their difference. However, if two things are really different, then switching things will change their difference. Then since the difference is lower, this 'permuted' condition will not be seen at the highest level in GSEA.

Here's the problem though. If there are say 8 conditions, by switching one pair at one time, we're really not CHANGING the difference too much, since one pair swap may not make too much difference (in our test, we're on average swapping 4 labels, so we're actually doing better).

My concern really is that if geneset A is so far different that swaps still maintain some difference, and the rest of the genesets are so far below that they never obtain even close by ES values, then geneset A will still come out on top, even though its ES value has gone far from the extreme that it started out with.

This above condition may happen when the ES scores are not somewhat continuously distributed in the non permuted calculation. In our downregulated set, the ES's come out to be somewhat continuous. However, in our upregulated set, the first two are abruptly different from the rest, and therefore they might have a tendency to still be enriched even though the permutes are 'diluting' their ES's very highly. Thus, in future work, it may be instructive to try to find a distribution of the ES values in each of the 1000 permutes and see where in that spectrum the non permuted ES falls. This may be a more reliable P value measure.

However, overall the P values should show somewhat what is happening in terms of whether the geneset if significant or not!

Now we look at row swaps and whether these are valid. What will a row swap do? A random row swap really doesn't DIFFERENTIATE whether the two test conditions are the same or different. The row swap will always swap randomly. It will never rank the genes the same way (unless that happens by chance). Thus, our VALUE or the rank order we obtained with the non permuted data is always permuted regardless of difference between our two test conditions. We always pass the permute test since we're now doing random orderings of our genes. It is not that this test is bad because it is too generous in passing all our genesets. It will make all the genesets have almost the same P values (if we truly rearrange the rows randomly)—assuming as seems reasonable that geneset length doesn't affect their ES scores. The test is bad because it simply doesn't let us pass the genesets when there is NO difference and pass them when there is a difference. In other words, the test is bad, because it doesn't let us discriminate between the two conditions. It completely bypasses the question of a test between the conditions. Thus, this test in randomly allocating the rows is useless.

Let's explore the idea of the P test further. What is the ranking of our genesets by enrichment score in the non permuted dataset really telling us? It is telling us that the top ranked geneset is the most significantly different between the stretch and control

conditions. The other genesets in turn are NOT significantly different between the stretch and control conditions. Thus, the top ranked geneset sets an upper bound on the P value in a sense. So if the P value for this geneset is extremely low—implying that this geneset passes the test for DIFFERENCE between control and stretch—then it is possible that genesets below it MAY also pass the P value test (if their P values are low enough—they can be higher than the P value of the top ranked geneset from the non permuted set if they are close enough in ES score to the top ranked non permuted dataset geneset). However, if the top ranked geneset in the nonpermuted dataset does NOT pass the P value test—implying that there really is no difference between the control and stretch conditions in this geneset—then really none of the genesets below it CAN pass, since the genesets were ranked based on a t statistic of the difference between the two conditions.

Thus the top rated geneset (in the non permuted dataset) sets the bar on the rest of the genesets. This top geneset decides if the rest of the geneset CAN be allowed to pass or CANNOT be allowed to pass depending on whether it itself passes the P value test!

We have to keep each column intact in our permutes, because each column is one condition. Essentially, we've created 8 pseudo-replicates in each condition (Control and Stretch). The reason we call these replicates 'pseudo' is that each of the samplings occur at different time points. The samplings are therefore not repeats at the same time and conditions.

Permutation testing implementation

Here's how the permutation test was implemented (perl was used to generate the whole code—see the Appendix):

Let's say our data matrix looks like the hypothetical example shown next:

1. Generate a random number for each stretch column (shown in the top row

	0.4	0.1	0.3	0.6	0.2	0.8	0.7	0.9
Stretch	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6	Time 7	Time 8
Gene1	S11	S12	S13	S14	S15	S16	S17	S18
Gene2	S21	S22	S23	S24	S25	S26	S27	S21
. . .								
GeneN	SN1	SN2	SN3	SN4	SN5	SN6	SN7	SN8

2. Sort the random numbers largest to smallest (keeping the columns associated with their corresponding random number).

	0.9	0.8	0.7	0.6	0.4	0.3	0.2	0.1
Stretch	Time 8	Time 6	Time 7	Time 4	Time 1	Time 3	Time 5	Time 2
Gene1	S18	S16	S17	S14	S11	S13	S15	S12
Gene2	S21	S26	S27	S24	S21	S23	S25	S22
. . .								
GeneN	SN8	SN6	SN7	SN4	SN1	SN3	SN5	SN2

3. Next generate random numbers for each control column.

	0.3	0.6	0.1	0.5	0.4	0.2	0.9	0.7
Control	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6	Time 7	Time 8
Gene1	C11	C12	C13	C14	C15	C16	C17	C18
Gene2	C21	C22	C23	C24	C25	C26	C27	C21
. . .								
GeneN	CN1	CN2	CN3	CN4	CN5	CN6	CN7	CN8

4. Swap with the newly paired stretch column (in grey above) only if the random number is larger than 0.5 (we're generating numbers between 0 and 1).
5. After the swap we obtain the following stretch and control matrices:

Final stretch matrix

Stretch	Time 8	Time 2	Time 7	Time 4	Time 1	Time 3	Time 7	Time 8
Gene1	S18	C12	S17	S14	S11	S13	C17	C18
Gene2	S21	C22	S27	S24	S21	S23	C27	C21
. . .								
GeneN	SN8	CN2	SN7	SN4	SN1	SN3	CN7	CN8

Final control matrix

Control	Time 1	Time 6	Time 3	Time 4	Time 5	Time 6	Time 5	Time 2
Gene1	C11	S16	C13	C14	C15	C16	S15	S12
Gene2	C21	S26	C23	C24	C25	C26	S25	S22
. . .								
GeneN	CN1	SN6	CN3	CN4	CN5	CN6	SN5	SN2

The above then gives us a permute (where we've switched classes—the classes being control and stretch) randomly. For each permute, we go through and calculate the enrichment score (as we discussed above) for each of the 173 gene sets that we have. Finally, we find the gene set with the highest enrichment score. This gene set then is counted once as having the highest enrichment score.

We run the permutation test 1000 times, each time findout the gene set with the maximum enrichment score (in our code that is given in the Appendix we implemented this by 'pushing' the gene set into a vector each time it obtained the highest enrichment score, and then counting the instance of each gene set in this vector at the end of the 1000 permutation tests).

We'll present the results of the gene set enrichment analysis for each of the three ways we performed the GSEA (up regulated only, down regulated only, and combined up and down regulated later). Next we'll illustrate how we attempted to find the 'right' time points between two time points (2 hour and 24 hour) that were inadvertently switched by the microarray facility. We were unsuccessful in determing which time point is which.

Switched time points

The two hour stretch and 24 hour stretch time points were inadvertently switched by the lab that handles the microarray hybridization process. The process was attempted to try to deduce which of the two (stretched) time points is which.

Saban et al [59] studied the time course gene expression of urinary tract inflammation in response to the placement of LPS in a mouse model. They built up 15 clusters of different genes that were differentially regulated over different time periods. Clusters 7, 9, and 10 were mentioned as being upregulated at earlier time points peaking within 4 hours and then declining by 24 hours. To see if these three clusters could unambiguously tell us which of the two scrambled time points is which, we assumed that one of the time points is the 2 hour time point and the other is the 24 hour time point. We then took ratios of the two with their respective controls (which are not scrambled). Next we made the opposite assumption and took two new ratios.

Next we checked for 39 genes that seemed to have a corresponding probe set ID at the affymetrix website. Unfortunately, we found an almost even split of 19 to 17 between the two assumptions (the reason these two numbers don't add up to 39 is that at times both assumptions led to a concurrent down or upregulation). The two numbers 19 vs 17 almost make it seem like the genes looked at are really not important in the time series and that the expression pattern is random with respect to deciphering the two time points.

It could just be that inflammatory genes are not really being over or under expressed with stretch (we will test this rigorously by applying GSEA by treating each cluster from the paper as a gene set). For now, therefore, we will discard these two time points and find clusters of genes by using paired GSEA, PCA, and SOM. The clusters that will be obtained at earlier (versus those obtained later) will then be used to find out if the two time points can be correctly placed unambiguously.

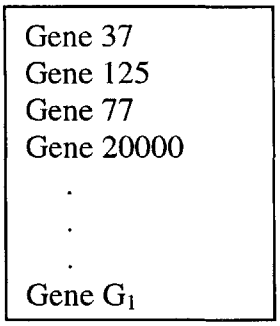
Creating the GSEA code without using high dimensional matrices

Initially the code was constructed by trying to build a set of matrices that added on columns each time a different computation was performed. Thus, each new computation had to carry all the previously loaded data into the next run of the algorithm. This worked when we calculated only one enrichment score. However, this technique was not very useful in performing 1000 permutations and at the same time creating a 1000 dimensional matrix each 2D section of which had 31099 by approximately 180 entries (each gene set was allocated a column). The memory soon ran out!

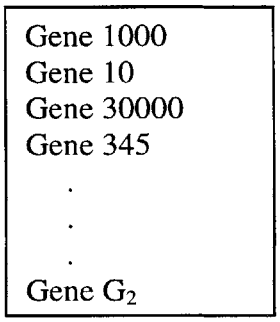
A much sparser solution was obtained by recognizing that first we didn't have to match each gene set with the reordered array with each new ordering. The two Kolmogorov statistic values are only dependent on the size of the group (the gene set) and the total number of genes in our analysis. Both of these stay the same for our whole analysis (the only way the total number of genes will change is if the chip is changed.

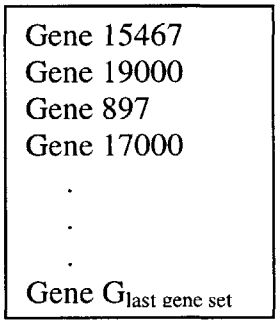
To see how we conducted the sparser solution, let's see how the genesets were loaded into Perl. We'll then study how these gene set files were transformed so that they had the Kolmogorov statistic on each line corresponding to the
We start out with text files, each text file having the list of probesets that make up one gene set. One extra text file stores the names of all the text files with the gene set names, one name per line. Let's see this graphically:

Geneset1.txt (say—in our system, we named the file based on the geneset name—for eg. angiogenesis.txt). Note that the size of a geneset given by G with the corresponding subscript is not constant across genesets. The made up gene numbers are given for each gene set.



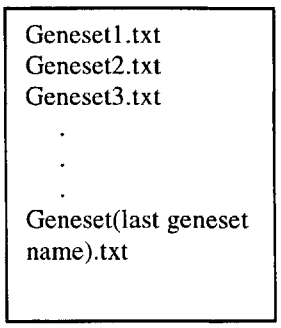
Geneset2.txt (this could for example be 'Response_to_mechanical_stimulus.txt')





After we created all the geneset text files, we created one text file called 'gsetfiles.txt'. This file had a list of all the gene set files, shown illustratively below:

gsetfiles.txt



Whenever we wanted to load the geneset files, we would first open a file handler in perl that would load the gsetfiles.txt and then we'd open filehandlers to each of the files listed in the gsetfiles.txt to access all the gene set text files. We'd then load each gene set gene into a suitable array to operate on it.

Early on, it was recognized that given that our gene sets are by definition heterogeneous both in terms of the numbers of genes they contain and the types of genes they contain, it would be computationally much too intensive to sort the the genes by the ranking metric (the t statistic squared as we used it or the difference metric that Mootha used) and then look for the corresponding entries that match with the members in each gene set. This step was made tractable by doing the following. Each gene (or probeset label) was replaced by an index running from 0 to 31098 (for a total of 31099 genes or probesets). For each geneset, we created a new file that had the appropriate Kolmogorov statistic in each row replacing the gene name. Let's look at this graphically:

Before we show the figures, it is important to remember that the gene set files to start with only have G_1, G_2, ..., G_n entries where the subscript refers to different genesets. So for example, while gene set 1 (G_1) may have 20 genes (one gene per row in the text file), gene set 2 may have 14 entries, gene set 34 may have 140 entries and so on. However, after we do the transformation, each gene set will have exactly 31099 entries. To make our illustration more concrete, let's use made up numbers. Let's say we have:

1. A total of 10 genes or probesets in our analysis, and let's say we have
2. Three gene sets, with
    a) gene set 1 having 2 members (genes 2, 9—we can't have duplicates),
    b) gene set 2 having 7 members (genes 7, 4, 9, 8, 10, 1, 5), and
    c) gene set 3 having 5 members (4, 2, 6, 1, 7).

We're now ready to see this graphically (see Figure on next page):

Generating the transformed Gene_set.txt files (for example, angiogenesis.txt , hypoxia.txt etc)

Gene_set_1.txt text file before transformation

Array of all genes to see where Gene_set_1 has correspondence

Call this file expression1.txt

Gene9
Gene2

Index (location) of gene	Name of gene
0	Gene1
1	Gene2
2	Gene3
3	Gene4
4	Gene5
5	Gene6
6	Gene7
7	Gene8
8	Gene9
9	Gene10

Compute,

$X = \sqrt{\frac{N-G}{G}} =$

$\sqrt{\frac{10-2}{2}} = 2$ when a member of Gene_set_1.txt matches the total gene array and place this value at locations 1 and 9. At all other locations, where our Gene_set_1.txt has no hit on the 'array of all genes' list, we place the alternate Kolmogorov statistic value given by,

$XN = -\sqrt{\frac{G}{N-G}} =$

$-\sqrt{\frac{2}{10-2}} = -0.5$

New Gene_set_1.txt file, with -0.5 at locations (corresponding to 'array of all genes' list) where there is no hit (shown in red), and 2 where there is a hit (shown in blue). Note that the transformed file has 10 entries

-0.5
2
-0.5
-0.5
-0.5
-0.5
-0.5
-0.5
2
-0.5

After generating the transformed file (which was named the same as the original file), we are ready for the analysis where we compute the permutations.

So let's start with our expression file (which we named expression1.txt) given above. Now let's add a column next to this file with hypothetical t statistic squared values:

Expression1. txt

Index (location) of gene	Name of gene	t statistic squared
0	Gene1	0.5
1	Gene2	0.4
2	Gene3	0.9
3	Gene4	0.1
4	Gene5	0.6
5	Gene6	0.7
6	Gene7	0.2
7	Gene8	0.95
8	Gene9	0.3
9	Gene10	0.8

Now we load our expression1.txt file into perl and reorder it based on the t statistic squared values. Rows stay together. Our index is thus shuffled as shown next:

Let's say we call the new array reordered array. To find the running sum, all we do is reference the gene_set.txt file through the index. Remember also that the new reordered array has its own index that starts at zero and goes to 9 (shown on left).

Reordered Array index	Index (location) of gene	Name of gene	t statistic squared
0	7	Gene8	0.95
1	2	Gene3	0.9
2	9	Gene10	0.8
3	5	Gene6	0.7
4	4	Gene5	0.6
5	0	Gene1	0.5
6	1	Gene2	0.4
7	8	Gene9	0.3
8	6	Gene7	0.2
9	3	Gene4	0.1

Reordered_array column 0

Thus, for example, if we have a 10 times 'for loop' to run through our reordered array (and let's say we load our (transformed) gene_set file into an array called gene_set_array), we can reference the first entry by performing 'gene_set_array[Reordered_array[0]]' to find the first running sum entry and so on.

The rest of the code (which is shown in the Appendix) was also made sparse by calling a subroutine that found the gene set with the highest enrichment value. This gene set name was then the only 'value' that was output back from the subroutine to our main program. This way our memory usage was kept to a bare minimum.

Self-organizing maps: A way to test validity of the gene set enrichment analysis that we performed

The self-organizing map was developed by Teuven Kohonen [60]. It involves finding closely matching patterns in a dataset. Essentially the method starts out by seeding with random vectors at certain locations and then inputting the data vectors at each of the locations. The smallest distance between the seeded vector and the dataset vector within a certain 'circular' range then gives us a new seeding vector that replaces the old vector (the new vector is the data set vector that gave us the smallest distance to the random starting vector). We can reseed the system with the data set vectors and continue the iteration until we converge.

The SOM then can classify data into classes based on the similarity of the vectors in each class.

GenePattern software from the Broad institute website [61] was used to find the most important genes that were up or down regulated. Although our data is noisy and therefore these genes are probably false positives in some sense, it is still useful to use these gene groups (one up regulated and one down regulated) to validate our gene set enrichment analysis because the two sets we obtain are essentially 'trained' on our data, but are obtained independently of our use of the GSEA.

We found that the up regulated genes from SOM were indeed enriched at the top of the up regulated GSEA, and that the down regulated genes are enriched at the top of the down regulated GSEA. Further, both these gene groups (or 'gene sets) came to the top in our combined analysis validating our analysis.

Now that we've discussed GSEA (including the permutation testing), let's look at some of the results that were obtained.

Results (GSEA)

First let's look at the combined up and down regulated GSEA

Combined up regulated and down regulated GSEA

The following table shows the enrichment scores for the first ten genesets along with their P values. The Appendix has the whole table with the enrichment scores along with the three different P value iterations for each gene set.

Gene Set	Enrichment Score	P value 1st iter.	P value 2nd iter.	P value 3rd iter.
hypoxia.txt	321.0	0.001	0.003	0.007
response_to_mechanical_stimulus.txt	310.3	0.014	0.015	0.016
c3_U133_probes.txt	308.1	0.018	0.018	0.018
MAP00340_Histidine_metabolism.txt	304.6	0.003	0.004	0.004
c27_U133_probes.txt	301.7	0.016	0.016	0.017
MAP00280_Valine_leucine_and_isoleucine_degradation.txt	283.6	0.008	0.011	0.01
c23_U133_probes.txt	278.6	0.013	0.01	0.012
c6_U133_probes.txt	254.7	0.01	0.017	0.014
mitochondr_HG-U133A_probes.txt	252.7	0.005	0.01	0.009
MAP00140_C21_Steroid_hormone_metabolism.txt	240.8	0.002	0.006	0

We see that hypoxia comes to the top in our combined analysis, and we see that its P value in the three iterations are very small, while we note that response_to_mechanical_stimulus finds itself second highest. Although this is reassuring, we should note that its P value is a bit high, telling us that it may have come close to the top as a chance event in our analysis.

Up regulated GSEA

The following table shows the results from the up regulated GSEA

Gene Set	Enrichment Score	P value
response_to_mechanical_stimulus.txt	560.2	0.085
c26_U133_probes.txt	415.6	0.014
c6_U133_probes.txt	343.6	0.023
c3_U133_probes.txt	336.8	0.003
Hum_Fb_Serum_EarlyTF.txt	335.1	0
OXPHOS_HG-U133A_probes.txt	328.9	0.131
c27_U133_probes.txt	289.0	0
mitochondr_HG-U133A_probes.txt	283.7	0.018
MAP00710_Carbon_fixation.txt	274.4	0.004
c14_U133_probes.txt	265.8	0.002

The high P value obtained by response to mechanical stimulus means that this gene set is not significant in the up regulated analysis

Down regulated GSEA

The following table shows the results of running the down regulated GSEA

Gene Set	ES	P value
MAP00280_Valine_leucine_and_isoleucine_degradation.txt	346.6	0.008
MAP00340_Histidine_metabolism.txt	288.9	0.005
MAP00310_Lysine_degradation.txt	284.6	0
MAP00380_Tryptophan_metabolism.txt	274.2	0.002
MAP00380_Tryptophan_metabolism~.txt	274.2	0
MAP00632_Benzoate_degradation.txt	270.5	0.01
hypoxia.txt	**263.3**	**0**
MAP00562_Inositol_phosphate_metabolism.txt	261.8	0.004
MAP00052_Galactose_metabolism.txt	242.5	0
MAP00521_Streptomycin_biosynthesis.txt	239.2	0.001

As above the full table is in the Appendix. We see here that the highest gene set has a low P value and (lower than 0.05) therefore is significant. We also note that hypoxia is at number 7. However, we see that Mootha's gene sets had a duplicate (numbers 4 and 5 above). Therefore, Hypoxia is actually at number 6, quite high and this points to the fact that hypoxia is more downregulated than upregulated. We'll discuss hypoxia at more length in the discussion to follow.

Discussion

The definition of hypoxia from [62] is:
"Reduction of oxygen supply to tissue below physiological levels despite adequate perfusion of the tissue by blood."

We found that in our combined gene set analysis, hypoxia came to the top, while in the down regulated analysis, hypoxia was quite close to the top. Many investigators have looked at the effects of hypoxia on cells grown in vitro. For example, Lu et al looked at rat pulmonary artery smooth muscle cells grown in vitro and showed increase in number of cells and increase in collagen synthesis under hypoxic conditions [63].

An in vitro study by An et al [64] studied the effects of hypoxia on rat pulmonary endothelial cells. The authors found increases in cell stiffness and traction forces exerted by cell on its substrate.

Another study by [65] looked at VEGF production by oral and epidermal keratinocytes in vitro. Much higher levels of VEGF protein and mRNA were reported by the authors in epidermal keratinocytes than in oral keratinocytes after 18 hrs of hypoxia.

Sahota et al studied "a number of simple strategies to improve angiogenesis/vascularization using a tissue-engineered model of skin to which small vessel human dermal microvascular endothelial cells were added". The variables studied were "cell type, angiogenic growth factors, the influence of keratinocytes and fibroblasts, mechanical penetration of the human dermis, the site of endothelial cell addition, and the influence of hypoxia"..."Two factors were identified that significantly enhanced endothelial cell penetration into the dermis: **hypoxia** [bold added for emphasis] and the site of endothelial cell addition." [66].

A common thread among the above cited studies is that in vitro, hypoxia induces pulmonary artery smooth muscle cells to become more proliferative, it induces

pulmonary endothelial cells to exert higher traction forces, that it increases angiogenesis and vacsularization of skin muscle cells, and that it induces VEGF expression in keratinocytes.

Hypoxia inducible factor (HIF) is a transcriptional factor that turns on genes that respond to hypoxia [67]. Kondo et al [67] show that HIF may play an important role in bladder and renal cancer cell proliferation and angiogenesis. Schlueter et al [68] state that tumor progression causes hypoxia conditions that upregulate genes that induce vascular proliferation.

It would seem reasonable that when we stretch rat ears, that cell proliferation would occur. Further, work at the Orgill lab has shown that rat ear stretch leads to growth in vasculature. However, the above lines of evidence which are extracted from in vitro studies seem to show that hypoxia is good for cell growth. On the other hand, hypoxia genes also are turned on in vivo under tumorigeneses and cancer like conditions.

Thus, it could be that there are separate pathways involved in neovascularization in tumorigenesis versus cell stretch and that the pathway involved in cell stretch is the one that suppresses hypoxia genes in our system. Further, in vivo effects because they are in balance with the rest of the organism are not extractable from in vitro systems. Thus, although hypoxia may confer certain benefits (although the metric of studying these benefits may themselves be uni-dimensional—for example, is cell growth always good? It certainly isn't when there is a tumor in the body) in vitro, it may do the reverse in vivo.

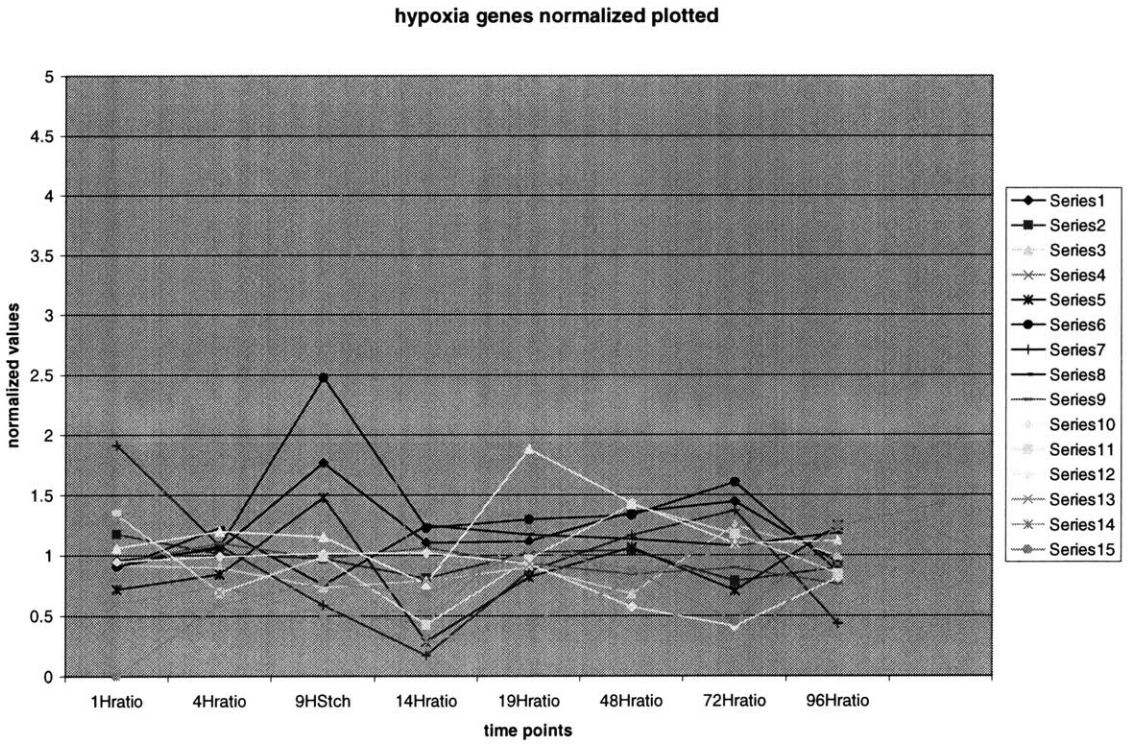
To understand the hypoxia geneset further, a time series analysis of hypoxia was conducted. Next, the hypoxia geneset was dissected into upregulated and downregulated components. A literature search combined with further runs on newer genesets created based on the literature search revealed possible mechanisms by which the hypoxia geneset may modulate tissue response to forces.

Let's look at the normalized time study of hypoxia first.

Time series study of 'hypoxia' looking for possible changes in regulation (down to up)

To gain an understanding of the analysis further, a time series study was conducted to see global patterns of gene expression. As can be seen in the graph below, there are tandem movements of genes each of which is given by the series labels (there are 15 genes in hypoxia). Four genes show remarkable in step movement at the earlier 9H time point. This seems to be the only interesting pattern that can be gleaned from this figure.

A normalization of stretch versus control was performed and then the hypoxia genes were plotted over time. There are 15 genes in hypoxia (each series in the chart below refers to a gene. The results are shown below.



To gain an understanding of which genes in hypoxia were upregulated in stretch versus control (over time), the t statistic was calculated. Its reordering gave the following results:

hypoxia genes (blue upregulated, red downregulated)

Probe Set ID	Gene Title	Gene Symbol	Chromosomal Location	Entrez Gene	t score
1368286_at	solute carrier family 2, (facilitated glucose transporter) member 8	Slc2a8	3p11	85256 Entrez gene	5.243583
1370080_at	**heme oxygenase (decycling) 1**	**Hmox1**	**19p11**	**24451 Entrez gene**	**4.126249**
1381936_at	calcium/calmodulin-dependent protein kinase II gamma	Camk2g	15p16	171140 Entrez gene	2.178111
1371719_at	bromodomain-containing 2	RGD:1303324	20p12	294276 Entrez gene	1.360505
1371289_at					-0.19267
1375650_at	bromodomain containing 4 (predicted)	RGD:1307282	7q11	362844 Entrez gene	-0.25929
1368322_at	**superoxide dismutase 3, extracellular**	**Sod3**	**14q11**	**25352 Entrez gene**	**-1.15325**
1369703_at	endothelial PAS domain protein 1	Epas1	6q12	29452 Entrez gene	-1.40849
1387605_at	caspase 12	Casp12	8q11	156117 Entrez gene	-1.41842
1387818_at	caspase 11	Casp4	8q11	114555 Entrez gene	-1.58963
1387667_at	**nitric oxide synthase 2, inducible**	**Nos2**	**10q24**	**24599 Entrez gene**	**-2.31805**
1373916_at	CDNA clone IMAGE:7303964, partial cds				-2.72021
1369186_at	caspase 1	Casp1	8q11	25166 Entrez gene	-2.77534
1374863_at					-2.99321
1369307_at	E1A binding protein p300	Ep300	7q34	170915 Entrez gene	-4.45035

Next, we're ready to collect inferences drawn from the literature search and put them together with further geneset enrichment analyses using newer genesets obtained from conclusions drawn from the literature search.

The response of tissue to stress is an important starting point for the literature search. A good example of this is reactive hyperemia and we begin with this.

Reactive Hyperemia

To gain an understanding of our system, it was decided to study the tissue response to mechanical stress, specifically the response of blood vessels to reactive hyperemia.

What is reactive hyperemia?

This is a response of blood vessels to a blockage in blood flow. For example, when a tourniquet is applied to the arm, blood flow is blocked. When the blood flow is reestablished, the increased blood flow [69] causes a rise in shear stress that then causes calcium influx in the endothelial cells, which then upregulates (nitric oxide) NO production, and this causes a relaxation of smooth muscle cells leading to an increase in the diameter of the vasculature (or vasodilation). It is believed that nitric oxide is upregulated during reactive hyperemia [69]. However, this mechanism of action is not established, and the role of NO in reactive hyperemia is largely unknown [70].

In our system, nitric oxide synthase II is being downregulated heavily. An attempt was made to explain this through pathway mechanisms. However, Zhao et al have shown that nitric oxide concentration does not increase during reactive hyperemia [71]. Wong et al have also obtained similar results in showing that in the cutaneous circulation (which itself is quite relevant to this thesis) NO synthase inhibition does not alter the reactive hyperemia response.

This leads to a possible conclusion that contrary to what is believed by many (that reactive hyperemia is mediated through NO production), it may be that NO mediates reactive hyperemia in certain locations and contexts (for example, Tagawa et al's and Wong et al's studies looked at the production of NO during exercise) while in other contexts it may not be very relevant.

It would be valid to conclude (we'll see why later through references) that NO production is a response to cell damage mediated by pathogens, tumor or other possible 'living' agents that need to be neutralized through the action of the innate immune system. Other pathways that may damage the cell (such as UV radiation) also can cause NO production [72]. However, cell stretch may be one pathway of cell injury that does not upregulate NO production.

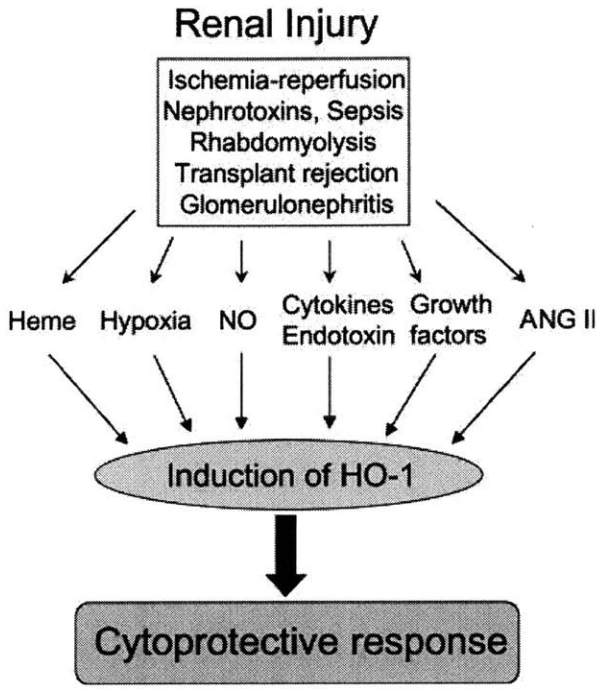
Heme Oxygenase I

In our system, however, Heme Oxygenase I, an enzyme with cell protective functions is highly upregulated, while nitric oxide synthase II is heavily downregulated. It is known that HO is upregulated by the production of NO [73]. However, HO itself is a negative regulator of NO. According to Turcano et al [74], induction of HO by hemin strongly inhibits NO production of LPS-activated macrophages.

Since HO is a negative regulator of NO, if HO is upregulated independently of NO production, then NO will be strongly downregulated. This then may explain why HO in our system is upregulated while NO is strongly downregulated. However, see the caveat at the end of this section.

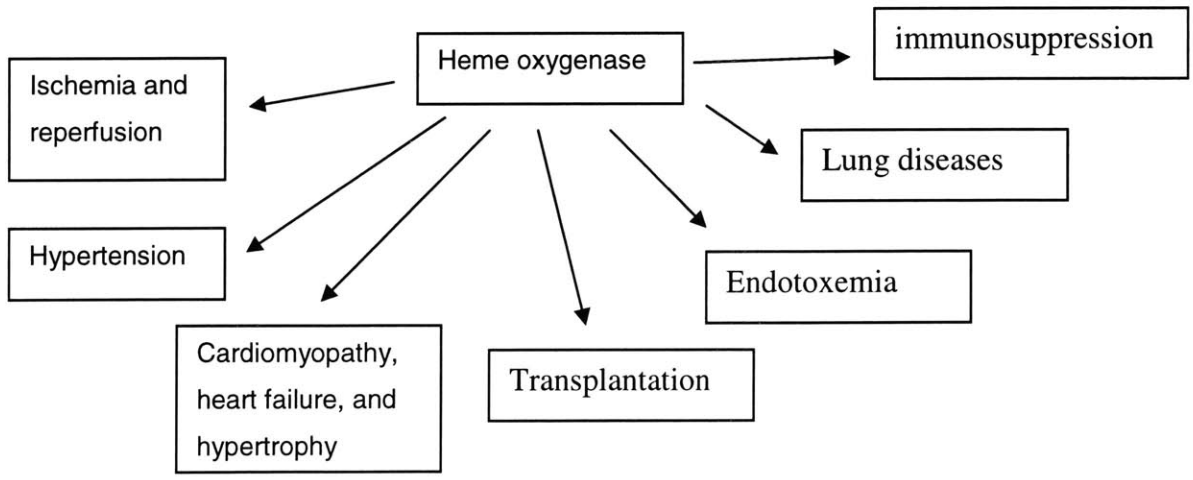
"HO-1, an inducible form of HO, is thought to act as an endogenous antioxidant defense mechanism" [75]. "Nitric Oxide Induces Heme Oxygenase-1 Gene Expression and Carbon Monoxide Production in Vascular Smooth Muscle Cells" [76]. "The Rat Heme Oxygenase-1 Gene Is Transcriptionally Induced via the Protein Kinase A Signaling

Pathway in Rat Hepatocyte Cultures" [77]. Many pathways lead to heme oxygenase I expression through renal injury [73]. The following figure is reproduced from [73]:



Thus, if mechanical stress can cause an upregulation in cytokines or growth factors which cause an overexpression in HO-1, then because HO-1 acts upon NO to downregulate it, we may get NO repression.

Heme Oxygenase itself has been implicated in various disorders [78]. The following figure is reproduced from [78]:



"High levels of HO-1 are frequently detected in various pathological states and generally in states of cellular oxidative stress." [79].

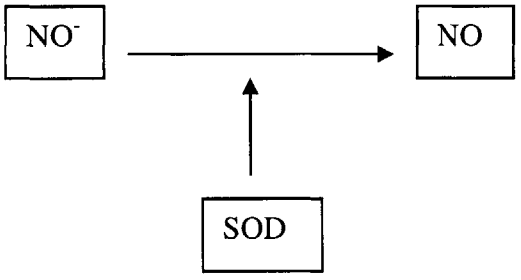
"HO-1 is a member of the heat shock protein family, which catalyzes the initial rate-limiting step of heme degradation to CO and biliverdin" [80].

Nitric Oxide

NO exhibits toxic effects, but the mechanisms are unclear [81]. Although mechanisms of the toxic effects exhibited by NO are unclear, various types of proposals [81] have been reported, including "monoADP ribosylation, S-nitrosylation of glyceraldehyde-3-phosphate dehydrogenase, inhibition of mitochondrial enzymes such as cis-aconitase, inhibition of mitochondrial electron transport chain, inhibition of ribonucleotide reductase, DNA damage, and activation of poly(ADP-ribose) synthetase."

NO is cytotoxic leading to apoptosis and is used in innate immunity and "tumor and pathogen killing" [82].

Murphy et al hypothesize that SOD [which is also part of the hypoxia geneset] promotes the production of NO from NO^- by accepting an electron from the latter [83]. This can be shown by the following pictorial:



In stretching cells, cell damage may lead to upregulation of heme oxygenase. Since there is no pathogen attack or tumor growth, NO production through SOD pathway isn't

initiated. Thus, since HO itself suppresses NO, the levels of NO are downregulated below baseline levels.

Keratinocytes' response to stretch

Yano et al have shown that in vitro, Keratinocytes respond to mechanical stretching by proliferating [84]. Zheng et al report "that FGF-2 and IGF-1 [which are growth factors] may regulate the proliferation step during [inner ear] hair cell development and regeneration." [85].

In this context it is interesting to note that "Platelet-derived growth factor suppresses and fibroblast growth factor enhances cytokine-induced production of nitric oxide by cultured smooth muscle cells." [86]

To test whether in our system PDGF is in fact upregulated while FGF may potentially not be so upregulated, queries were performed at "NetAffx" [87] for PDGF and FGF probesets corresponding to the RAE 230 2.0 rat chips. Gene sets were then constructed out of the probeset ID's obtained.

These genesets were then input into the UPregulated gene set analysis to see if they show high ES scores. Interestingly, the PDGF geneset obtained an ES of 152 (FGF had a score of 115). PDGF and FGF had ranks 49 and 70 out of 177 total genesets).

PDGF is a growth stimulant which is seen in monocyte supernatants [88]. **We can thus conclude that indeed cytokines that cause cell proliferation are upregulated in our system and that there is cell growth in our system.**

It is not clear whether there is tissue injury in the system

Next, we wanted to see if tissue injury cytokines may be upregulated (since nitric oxide is a mediator of cell defense and this is downregulated, we'd expect cytokines that feed into

nitric oxide synthase to be not so very upregulated if they are upregulated at all). This would be reasonable since really there is no foreign microbial or 'living' agent that needs to be protected against or neutralized.

Any injury that may potentially occur and that may cause free radical generation should be contained by factors that neutralize these possible free radicals—one might be NADPH. Glucose 6-phosphate dehydrogenase is the first enzyme in the pentose phosphate pathway and controls the production of NADPH (a "...reductant essential in many biosynthetic pathways" pg 551 [89]). Only one probeset was obtained on a search for G6PD. Its t squared statistic value is 0.7—not very high, but still positive.

Liu et al [90] mention that lactate dehydrogenase is not a very sensitive marker for cell injury by stretch. It was decided nevertheless to see if lactate dehydrogenase may possibly be upregulated in our system. Surprisingly, a lactate dehydrogenase geneset went straight to the bottom (of the UPregulated analysis) with an ES score of 0.9 (the highest ES score was OXPHOS_HG-U133A_probes with a score of 446.

Brickson et al [91] performed "a standardized single stretch injury ...[by subjecting the rabbits to acute stretch]... on the tibialis anterior (TA) muscle of 36 male New Zealand white rabbits ...[with] contralateral control limbs ...[undergoing] a sham surgery". Their results show an increase in the production of superoxide dismutase and Xanthine oxidase.[91]

Thus SOD and XO genesets were created and the ES scores obtained were 100 and 74 out of 177 genesets (not very high)—their ranks were 81 and 104 respectively. This result isn't very surprising, since we saw earlier that SOD modulates the conversion of NO^- to NO, and NO itself is downregulated. Thus, we need to find other markers of stretch injury (if indeed there is "injury" in our system). Collagen XII gene expression was found to be increased in an ex vivo model of rat bladder stretch injury[92].

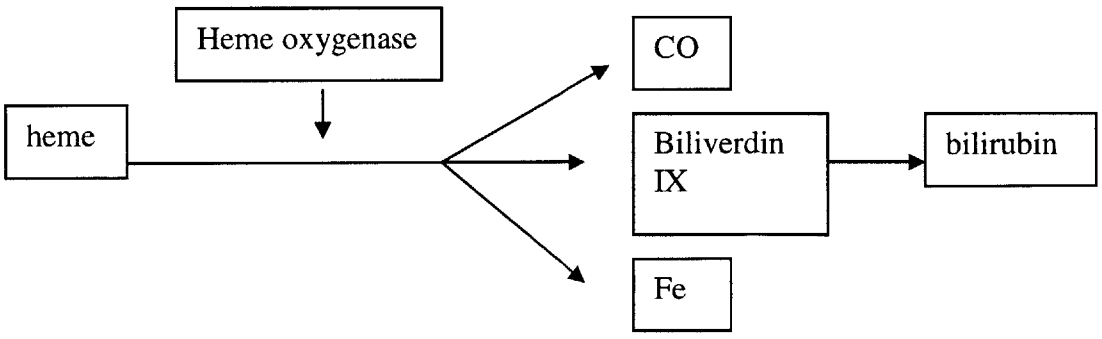
A search on NetAffx yielded only one probeset for "collagen XII". Although searching for one probeset may be prone to noise, its t statistic squared value is still useful because it is extracted from 8 paired time points and also because it still gives us some idea of its up or down regulation. We found that this probeset was slightly downregulated with a t squared statistic of -1 approximately.

Thus, at least the genes (or genesets) that we've looked at, seem to show that there is little cell injury. However, heme oxygenase is being heavily upregulated, and therefore there must be some damage that this enzyme is protecting against. For example, Otterbein et al [93] show that "Exogenous administration of heme oxygenase-1 by gene transfer provides protection against hyperoxia-induced lung injury". This paper interestingly also shows that in this system (where HO-1 gene was overexpressed in the rat lung through gene transfer), **HO-1 doesn't upregulate superoxide dismutase or L-ferritin or H-ferritin (this is exactly what we had predicted earlier, since in our system NO is suppressed and SOD upregulates NO).**

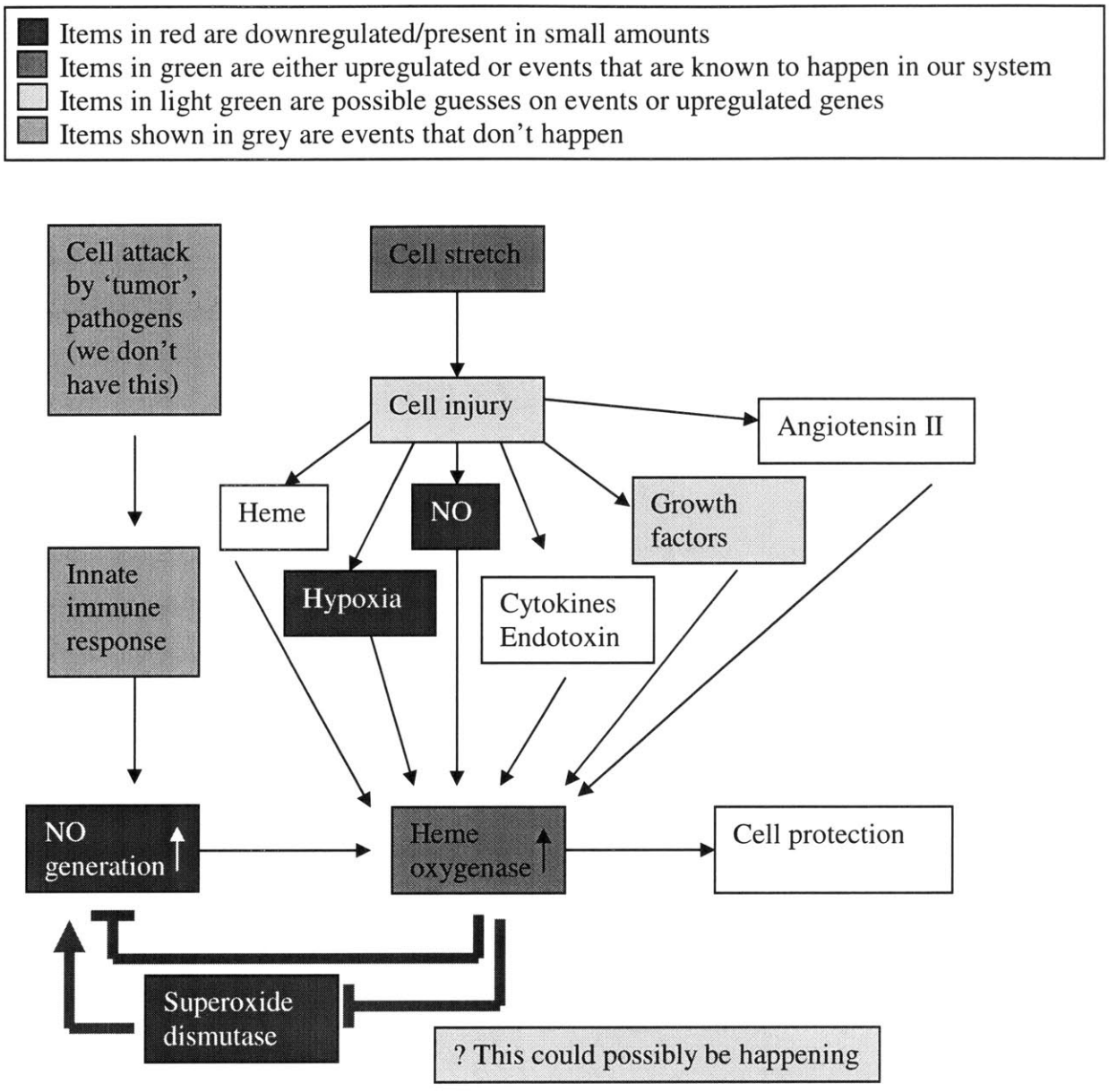
Before we deduce the pathway that may be at work, let's see what heme oxygenase does downstream.

Heme oxygenase downstream action

The following pathway shows that heme oxygenase modulates the transformation of heme to CO, biliverdin IXa (which is then converted to bilirubin), and iron [93], [94]



The pathway that may be at work



For comparison, the hypoxia gene set is given in the previous section (the items in our pathway are in bold and larger font).

Caveats in drawing the above pathway conclusion

A caveat that should be kept in mind is that gene set enrichment analysis damps the effects of noise by looking at the 'movement' of a group of genes in a system rather than the 'movement' of one gene itself. Thus, even though it is reasonable to conclude that hypoxia itself is in general the most regulated, drawing conclusions about each gene in the set may lead to erroneous conclusions. These conclusions may especially be the more erroneous if we try to draw comparisons between the regulations of genes both of which may move in the same direction but show very little difference in the t statistic squared scores.

Another caveat that must be given is that the t scores themselves while not small aren't the largest seen in the system. It is their aggregate value as seen through the Kolmogorov Smirnov statistic that lends the GSEA its strength and usefulness.

One mitigating and helpful factor is that in our analysis we had eight time points that were paired, and thus in essence the t statistic itself had a basis in 'strength in numbers' in predicting the 'movement' of genes (and not just genesets). This gives us some confidence in the strength of the results obtained.

To gain a further understanding of the system, the top 10 genes from each of the three analyses (combined, up and down regulated) were obtained.

The top 10 genes in the analysis were found (combined, the upregulated, and the downregulated).

COMBINED top 10 genes

	Probe Set ID	Gene Title	Gene Symbol	GO Biological Process Description	GO Molecular Function Description	GO Cellular Component Description	t score
1	1367680_at	acyl-Coenzyme A oxidase 1, palmitoyl	Acox1	electron transport fatty acid metabolism fatty acid beta-oxidation spermatogenesis	acyl-CoA oxidase activity oxidoreductase activity	mitochondrion peroxisome	**103.0628**
2	1397750_at						**91.64285**
3	1387122_at	pleiomorphic adenoma gene-like 1	Plagl1	regulation of cell cycle apoptosis	nucleic acid binding		**87.49114**
4	1389422_at						**63.66021**
5	1372900_at	phosphatidylinositol glycan, class T (predicted)	RGD:1307156				**60.63113**
6	1391406_at	Degenerative spermatocyte homolog (Drosophila)	Degs		oxidoreductase activity sphingolipid delta-4 desaturase activity	integral to membrane	**58.67817**

7	1395901_at	YY1 transcription factor	Yy1	regulation of transcription, DNA-dependent negative regulation of steroid metabolism	nucleic acid binding DNA binding transcription factor activity zinc ion binding specific transcriptional repressor activity	nucleus transcription factor complex	**56.17106**
8	1387956_s_at	chemokine-like factor 1	RGD:620708	neutrophil chemotaxis immune cell chemotaxis	cytokine activity	integral to membrane	**56.09695**
9	1371771_at	Transcribed locus, strongly similar to XP_233606.2 perlecan [Rattus norvegicus]					**91.64285**
10	1390494_at	ribosomal protein S24	Rps24	protein biosynthesis ribosome biogenesis	RNA binding structural constituent of ribosome	intracellular ribosome	**56.17106**

					cytosolic small ribosomal subunit (sensu Eukaryota)		

UP top 10 genes

	Probe Set ID	Gene Title	Gene Symbol	GO Biological Process Description	GO Molecular Function Description	GO Cellular Component Description	t score
1	1372900_at	phosphatidylinositol glycan, class T (predicted)	RGD:1307156				60.63113
2	1368130_at	aldehyde dehydrogenase family 3, member A1	Aldh3a1	metabolism	aldehyde dehydrogenase activity aldehyde dehydrogenase (NAD) activity aldehyde dehydrogenase [NAD(P)+] activity oxidoreductase activity		50.44276
3	1379849_at	THO complex 3 (predicted)	RGD:1311669				50.05069

4	1388402_at	similar to 2410001H17Rik protein (predicted)	RGD1305824_predicted				48.82479
5	1398912_at	mitochondrial ribosomal protein L9	mrpl9				42.68423
6	1387456_at	staufen, RNA binding protein, homolog 2 (Drosophila)	RGD:621479		double-stranded RNA binding	intracellular	39.19826
7	1388612_at	Similar to RIKEN cDNA 6030432N09					38.6693
8	1371134_at	ATPase, Ca++-sequestering	Atp2c1	cation transport calcium ion transport metabolism proton transport	magnesium ion binding catalytic activity calcium-transporting ATPase activity ATP binding	membrane integral to membrane	37.59454

				secretory pathway	ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism hydrolase activity hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances		
9	1382793_at	Transcribed locus					37.46606
10	1369092_at	sec22 homolog	Sec22l2				36.82897

DOWN top 10 genes

	Probe Set ID	Gene Title	Gene Symbol	GO Biological Process Description	GO Molecular Function Description	GO Cellular Component Description	t score
1	1367680_at	acyl-Coenzyme A oxidase 1, palmito	Acox1	electron transport fatty acid metabolism fatty acid beta-oxidation	acyl-CoA oxidase activity oxidoreductase activity	mitochondrion peroxisome	103.06281

		yl		spermatogenesis			
2	1397750_at						91.642852
3	1387122_at	pleiomorphic adenoma gene-like 1	Plagl1	regulation of cell cycle apoptosis	nucleic acid binding		87.491139
4	1389422_at						63.660214
5	1391406_at	Degenerative spermatocyte homolog (Drosophila)	Degs		oxidoreductase activity sphingolipid delta-4 desaturase activity	integral to membrane	58.678174
6	1395901_at	YY1 transcription factor	Yy1	regulation of transcription, DNA-dependent negative regulation of steroid metabolism	nucleic acid binding DNA binding transcription factor activity zinc ion binding specific transcriptional repressor activity	nucleus transcription factor complex	56.171056
7	1387956_s_at	chemokine-like factor 1	RGD:620708	neutrophil chemotaxis immune cell chemotaxis	cytokine activity	integral to membrane	56.09695
8	1371771_at	Transcribed locus, strongly similar to XP_233606.2 perlecan [Rattus norvegi					53.49104

		cus]					
9	1390494_at	ribosomal protein S24	Rps24	protein biosynthesis ribosome biogenesis	RNA binding structural constituent of ribosome	intracellular ribosome cytosolic small ribosomal subunit (sensu Eukaryota)	53.488477
10	1383150_at	Transcribed locus					48.369693

Next, the bootstrap technique of permutation testing (which uses replacement as the permutations are conducted) was performed to see if indeed as Good mentions [58], both the bootstrap and the permutation test give us comparable results.

Bootstrap analysis

The results of the bootstrap analysis were very similar to the P test of the permutation test. Two iteration results are given below:

Bootstrap results 1st iteration

geneset	P value
c2_U133_probes.txt	0.075
c7_U133_probes.txt	0.053
OXPHOS_HG-U133A_probes.txt	0.047
c18_U133_probes.txt	0.035
MAP00100_Sterol_biosynthesis.txt	0.023
c6_U133_probes.txt	0.023
c1_U133_probes.txt	0.019
MAP00900_Terpenoid_biosynthesis.txt	0.017
MAP00970_Aminoacyl_tRNA_biosynthesis.txt	0.017
MAP00280_Valine_leucine_and_isoleucine_degradation.txt	0.016
cluster15_LPS_mouse_urinary.txt	0.016
c3_U133_probes.txt	0.015

c5_U133_probes.txt	0.015
c33_U133_probes.txt	0.014
response_to_mechanical_stimulus.txt	0.013
c0_U133_probes.txt	0.012
c27_U133_probes.txt	0.012
c9_U133_probes.txt	0.012
MAP00430_Taurine_and_hypotaurine_metabolism.txt	0.011
c25_U133_probes.txt	0.011

Bootstrap results 2nd iteration

geneset	P value
c7_U133_probes.txt	0.068
c2_U133_probes.txt	0.052
OXPHOS_HG-U133A_probes.txt	0.05
c18_U133_probes.txt	0.032
c6_U133_probes.txt	0.025
c24_U133_probes.txt	0.023
MAP00100_Sterol_biosynthesis.txt	0.02
MAP00970_Aminoacyl_tRNA_biosynthesis.txt	0.02
c1_U133_probes.txt	0.018
c33_U133_probes.txt	0.016
c15_U133_probes.txt	0.015
cluster10_LPS_mouse_urinary.txt	0.015
response_to_mechanical_stimulus.txt	0.015
MAP00900_Terpenoid_biosynthesis.txt	0.014
c5_U133_probes.txt	0.014
c25_U133_probes.txt	0.012
c35_U133_probes.txt	0.012
c0_U133_probes.txt	0.011
cluster15_LPS_mouse_urinary.txt	0.011
inflammatory.txt	0.011

For comparison, here are the results of the Permutation test (only one iteration shown):

Permutation test

geneset	P value
c7_U133_probes.txt	0.068
c2_U133_probes.txt	0.062
OXPHOS_HG-U133A_probes.txt	0.033
c18_U133_probes.txt	0.032
MAP00100_Sterol_biosynthesis.txt	0.022
MAP00900_Terpenoid_biosynthesis.txt	0.019
c3_U133_probes.txt	0.018
human_mitoDB_6_2002_HG-U133A_probes.txt	0.018
c1_U133_probes.txt	0.016
c27_U133_probes.txt	0.016
MAP00562_Inositol_phosphate_metabolism.txt	0.014
TCA_HG-U133A_probes.txt	0.014
c0_U133_probes.txt	0.014

c24_U133_probes.txt	0.014
c9_U133_probes.txt	0.014
response_to_mechanical_stimulus.txt	0.014
c15_U133_probes.txt	0.013
c23_U133_probes.txt	0.013
c5_U133_probes.txt	0.013
MAP00561_Glycerolipid_metabolism.txt	0.012

Singular value decomposition

Any n x p matrix can be decomposed into a left orthonormal n x p matrix, a square p x p diagonal matrix of eigenvalues, and a right orthonormal p x p matrix called the 'eigenassay' [95]. The left eigenvector or eigengene coefficients tell us about the important variances in the genes (**however since our rank has to be either 8 or lower than 8, it would seem that we can at most have only 8 significant variances down each column**).

Obtaining the SVD on the dataset.

We have 8 time points control and 8 time points stretched. First we take the ratio of stretch to control to obtain the fold change in the expression profiles. Now we have the starting data matrix on which we'll perform the SVD.

Save the data as a text file (this can be done in excel). Each row of the text file contains the 8 time point normalized values for a single gene. In R, we save this file in the C:\Program Files\R\rw2010 directory (or whatever directory R reads from). To load this file into a matrix Z, we do the following:
Z<-matrix(scan("ratio.txt"),nrow=31099,byrow=T)
The above will load the file 'ratio.txt' which into a matrix with 31099 rows (that's the number of genes we're scanning in—technically 31099 is really the number of probesets that we're scanning in, not the number of genes).

To perform the SVD on the Z matrix, type:
decomposition<- svd(Z)
The results of the svd are then stored in the variable 'decomposition'
To see the U, S, and V matrices of the SVD, we can type
decomposition$u
decomposition$v
decomposition$d

To obtain each of the eight left eigenvectors (with 31099 components each), we can essentially read each column of the U matrix (given by decomposition$u) into a new vector. This is done by typing,

u1<-decomposition$u[,1]

Note that there is a space before the comma in the square brackets. This will load the first column into the vector u1. u2 would be given by

U2<-decomposition$u[,2]

and so on.

Now we load each of these eight vectors into eight separate text files by typing,

write(u1,ncol=1,"u1.txt")

In the above command, note that we need to specify ncol=1. Otherwise, 'R' will load the data into more than one column.

Next we created a two column file (say 'expression1.txt') with indices starting at 0 in the first column and the coefficients in the second column (the absolute values of the coefficients are used since the variation is captured in the absolute values—the sign being unimportant)s. We next ran the GSEA on this set.

Results

The top ten genesets for each left eigenvector are given below:

LEV=left eigenvector

LEV1

response_to_mechanical_stimulus.txt	1
MAP00361_gamma_Hexachlorocyclohexane_degradation.txt	**6**
c14_U133_probes.txt	1
MAP00430_Taurine_and_hypotaurine_metabolism.txt	**6**
Hum_Fb_Serum_EarlyTF.txt	2
MAP00480_Glutathione_metabolism.txt	3
MAP00150_Androgen_and_estrogen_metabolism.txt	**4**
c26_U133_probes.txt	1
MAP00940_Flavonoids_stilbene_and_lignin_biosynthesis.txt	**6**
OXPHOS_HG-U133A_probes.txt	1

LEV2

c24_U133_probes.txt	3
MAP00940_Flavonoids_stilbene_and_lignin_biosynthesis.txt	**6**
cluster11_LPS_mouse_urinary.txt	3
c29_U133_probes.txt	1
MAP00430_Taurine_and_hypotaurine_metabolism.txt	**6**
MAP00361_gamma_Hexachlorocyclohexane_degradation.txt	**6**
c2_U133_probes.txt	1
cluster15_LPS_mouse_urinary.txt	2
MAP00602_Blood_group_glycolipid_biosynthesis_neolact_series.txt	1
MAP00040_Pentose_and_glucuronate_interconversions.txt	3

LEV3

MAP00150_Androgen_and_estrogen_metabolism.txt	4
MAP00361_gamma_Hexachlorocyclohexane_degradation.txt	**6**
c31_U133_probes.txt	1
MAP00940_Flavonoids_stilbene_and_lignin_biosynthesis.txt	**6**
MAP00140_C21_Steroid_hormone_metabolism.txt	2
MAP00040_Pentose_and_glucuronate_interconversions.txt	3
c24_U133_probes.txt	3
perception_of_pain_sensory_transductuon_of_mechanical_stimulus.txt	1
sensory_perception_of_mechanical_stimulus.txt	1
cluster15_LPS_mouse_urinary.txt	2

LEV4

ECM_constituent_conferring_elasticity.txt	1
ECM_structural_constituent_conferring_compression_resistance.txt	1
FA_HG-U133A_probes.txt	1
GLUCO_HG-U133A_probes.txt	1
GLYCOGEN_HG-133A_probes.txt	1
GLYCOL_HG-U133A_probes.txt	2
GO_0005739_HG-U133A_probes.txt	1
Hum_Fb_Serum_Chol_Biosyn.txt	1
Hum_Fb_Serum_Coag_Hemo.txt	2
Hum_Fb_Serum_EarlyTF.txt	2

LEV5

MAP00051_Fructose_and_mannose_metabolism.txt	2
c24_U133_probes.txt	3
MAP00430_Taurine_and_hypotaurine_metabolism.txt	**6**
Hum_Fb_Serum_Coag_Hemo.txt	2
MAP00361_gamma_Hexachlorocyclohexane_degradation.txt	**6**
cluster11_LPS_mouse_urinary.txt	3
MAP00040_Pentose_and_glucuronate_interconversions.txt	3
MAP00480_Glutathione_metabolism.txt	3
GLYCOL_HG-U133A_probes.txt	2
c22_U133_probes.txt	2

LEV6

MAP00430_Taurine_and_hypotaurine_metabolism.txt	**6**
MAP00251_Glutamate_metabolism.txt	1
MAP00361_gamma_Hexachlorocyclohexane_degradation.txt	**6**
MAP00740_Riboflavin_metabolism.txt	1
MAP00512_O_Glycans_biosynthesis.txt	2
MAP00960_Alkaloid_biosynthesis_II.txt	1
MAP00940_Flavonoids_stilbene_and_lignin_biosynthesis.txt	**6**
MAP00051_Fructose_and_mannose_metabolism.txt	2
c22_U133_probes.txt	2
cluster11_LPS_mouse_urinary.txt	3

LEV7

MAP00361_gamma_Hexachlorocyclohexane_degradation.txt	**6**
MAP00940_Flavonoids_stilbene_and_lignin_biosynthesis.txt	**6**
cluster4_LPS_mouse_urinary.txt	1
MAP00150_Androgen_and_estrogen_metabolism.txt	4
MAP00512_O_Glycans_biosynthesis.txt	2
MAP00430_Taurine_and_hypotaurine_metabolism.txt	**6**
MAP00140_C21_Steroid_hormone_metabolism.txt	2
MAP00533_Keratan_sulfate_biosynthesis.txt	1
inflammatory.txt	1
MAP00472_D_Arginine_and_D_ornithine_metabolism.txt	2

LEV8

cluster14_LPS_mouse_urinary.txt	1
inflammatory_mechanical.txt	1
MAP00430_Taurine_and_hypotaurine_metabolism.txt	**6**
MAP00940_Flavonoids_stilbene_and_lignin_biosynthesis.txt	**6**
MAP00150_Androgen_and_estrogen_metabolism.txt	**4**
cluster2_LPS_mouse_urinary.txt	1
MAP00580_Phospholipid_degradation.txt	1
MAP00710_Carbon_fixation.txt	1
MAP00472_D_Arginine_and_D_ornithine_metabolism.txt	2
MAP00480_Glutathione_metabolism.txt	3

If the above analysis has validity, it does show as expected that 'response to mechanical stimulus' is the most important gene set as far as time series are concerned. This analysis has thus extracted the most important geneset from one time point to another. The analysis we did where we ordered the genes with the T statistic squared really looked at the difference between stretch and control. Here we are no longer saying that 'response to mechanical stimulus' differentiates the stretch versus the control conditions. We are saying that this geneset is important in going from one time point to another (it may however also have importance in looking at stretch versus control—as it should, but that information isn't obtainable from the SVD analysis. We have to do the T statistic GSEA to see that separately.

Chapter 5

Finite element analysis using Biphasic wound model

To replicate the wound more faithfully, and also to address commentary that accompanied our solid finite element publication, a biphasic finite element model was constructed.

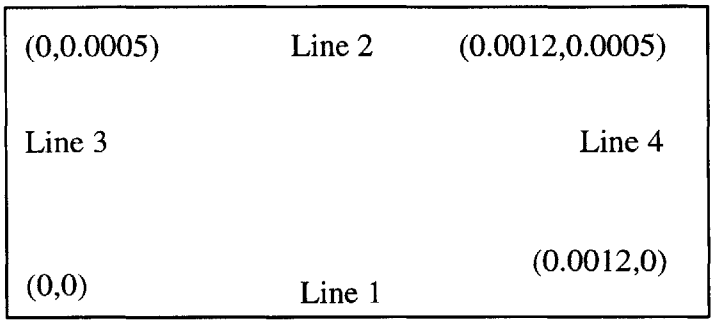
The Finite element model

As before, ADINA (ADINA inc, Watertown MA) was used to conduct the analysis. A 2D model was created using the FSI (or fluid structure interfact) formulation in ADINA. 4 node elements were used (9 node elements were unsuccessful in our model). To use the FSI formulation, two different models have to be created, one model in the ADINA module (which pertains to solid conditions) and one in the ADINA-F (F referring to fluid) module. Each module has the FSI condition on the menu applied (or checked). Let's look at the details of the model building step in each module

ADINA (solid)

1) Under control→degrees of freedom, leave the checks only on the Y-translation and Z-translation options
2) Next control → fluid structure interaction → check both 'fluid structure interaction' and 'includes porous coupling'
3) Geometry → points → enter the points (X1 stands for the 'X' coordinate. In the 2D system, we really only use the X2 and X3 coordinates or the Y and Z coordinates. Therefore for our analysis, we leave the X1 coordinate blank. Also, leaving the 'system'—which stands for coordinate system blank implies that we use the default 'zero' coordinate system.
4) Geometry → using the points we then create → surfaces

5) The next picture shows our model with the coordinate point locations for the solid model (remember, we're creating only the solid section for now). Also, as opposed to the solid only analysis where we actually modeled the struts, in this case, we're only modeling the intra-strut space.



6) Model → porous media properties are initially greyed out, but this item needs to be set (even though the values we specify don't seem to make a difference). We first need to specify the element group and then specify 'porous media properties'. First however, let's specify the materials, boundary conditions and loads.
7) Model → material → elastic → isotropic → add (new material) → enter the values of the Young's modulus, the Poisson's ratio, and the density (although the density isn't really important since this is not a dynamic analysis (where transient terms are important).
8) Model → Boundary conditions → FSI boundary → add (new FSI) → apply to line at bottom (shown as line 1 above). Now, it is important to note that each non-continuous FSI condition has to be applied with a new FSI number. Therefore, we have to click on 'add' again and then apply another FSI boundary on line 2 (above).
9) Model → Boundary conditions → define fixity → add (call this fixity yfixed) → Fixed degrees of freedom → check the y translation box (leave the z translation unchecked). Apply this fixity next to lines 3 and 4 above.
10) (Applying the above created fixity) Model → Boundary conditions → Apply fixity → Apply to: Lines, (choose lines 3 and 4 above), and under 'apply fixity' column,

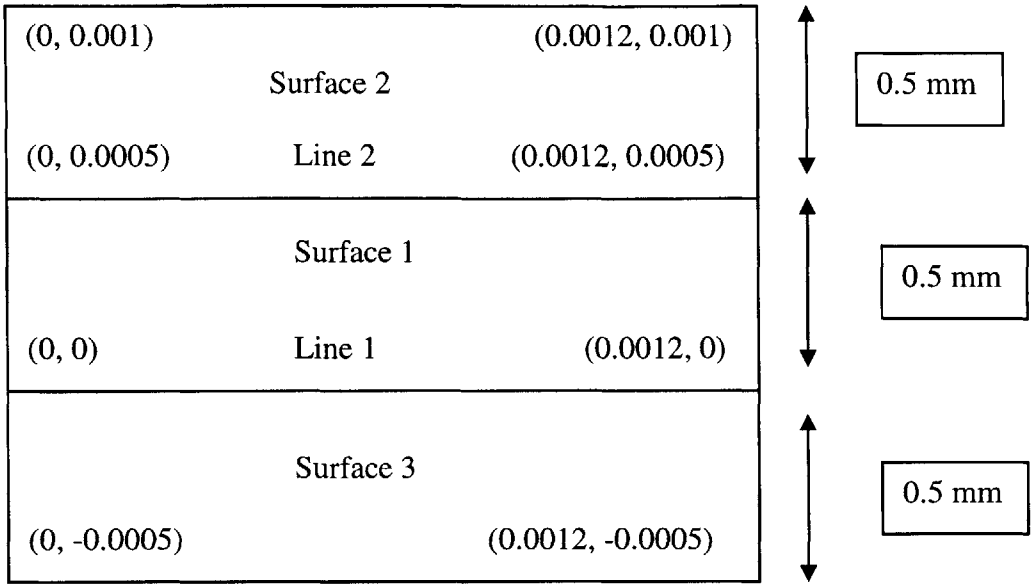
choose from the drop down option (this option becomes visible when we click on the rows under this column) yfixed.

11) Model → Loading → Apply → Load type: displacement, Apply to (leave as point), Load number: click 'define' → Add → Prescribed values of translation → X:0, Y:0, Z:0 (we could have also defined fixity rather than zero displacement loading to get the same result on this loading/boundary condition).
12) Now, we go back and apply 'porous media property'—this option is no longer grey; Model → materials → porous media property → don't change anything, click OK.
13) Meshing → element groups → add (new element group) → type: 2D solid, Default material (leave as 1), element sub-type: plane strain, element option: porous media.
14) Meshing → mesh density → surface → Number of subdivisions: u: 20, v: 10 (we may have to swap these depending on which way the surface ended up being defined. We essentially want the larger number of divisions on lines 1 and 2 above).
15) Meshing → create mesh → surface → element group (leave this as 1), Nodal specification: Nodes per element: change to 4 from 9. Surface: enter 1. Click OK.
16) Solution → Data File/Run → **uncheck the 'Run ADINA' option** (because we only want to save the data file rather than running ADINA—we'll do the same under the fluid option, and then we'll run the two files we generate together under FSI).
17) After running the data file creation above, click on SAVE to save the file. Then go to File → new. We're ready for the fluid analysis.

Next we look at ADINA-F model creation.

ADINA-F model creation

1) Choose ADINA-F module, click on FSI
2) Model → flow assumptions → Flow dimension: 2D, uncheck 'includes heat transfer', 'Fluid Structure interaction': check FSI, check 'includes porous coupling'.
3) We create the fluid model. Geometry → points → enter the points as shown in the pictorial below:



4) Now we'll define two different fluid materials, the material in the center being porous fluid, and the material on either side being a regular 'non-porous' fluid. First, let's create the surfaces using the vertices we listed above.
5) Model → materials → constant → add (this will be listed or added as material 1) → viscosity: 0.015, density: 1000
6) Model → materials → porous → add (this will be listed or added as material 2) → fluid: density: 1000, permeability: X: 10e-16, Y: 10e-16, Z: 10e-16, Porosity: 0.5, viscosity: 0.015.
7) Model → Usual Boundary conditions/Loads → Zero values → add → call this 'noYvelocity' → degrees of freedom with zero values → check 'Y-velocity' (we will apply this 'zero value' boundary condition to lines.
8) Model → Usual Boundary conditions/Loads → apply zero values → apply to: lines, under 'line #' pick the lines, under zero values (when we click in this column the drop down list with all zero value options becomes available) pick noYvelocity for each line number. OK
9) Model → special boundary conditions → add → type: fluid-structure interface, slip condition: no, apply to: line 1. NOW click ADD and then repeat on line 2 in the figure above (same reasoning as before—if we have two discontinuous FSI boundaries, we need to create separate FSI interaction numbers).

10) Model → Usual Boundary conditions/Loads → apply → Load type: Normal Traction, apply to: line, load number → define → add → enter magnitude (say 10000), choose line at bottom.

11) Meshing → element group → material 1 → element sub-type: planar, 2D fluid (default)

12) Meshing → element group → material 2 → element sub-type: planar, 2D fluid (default)

13) Meshing → mesh density → 4 (down), 5 (across) (all surfaces)

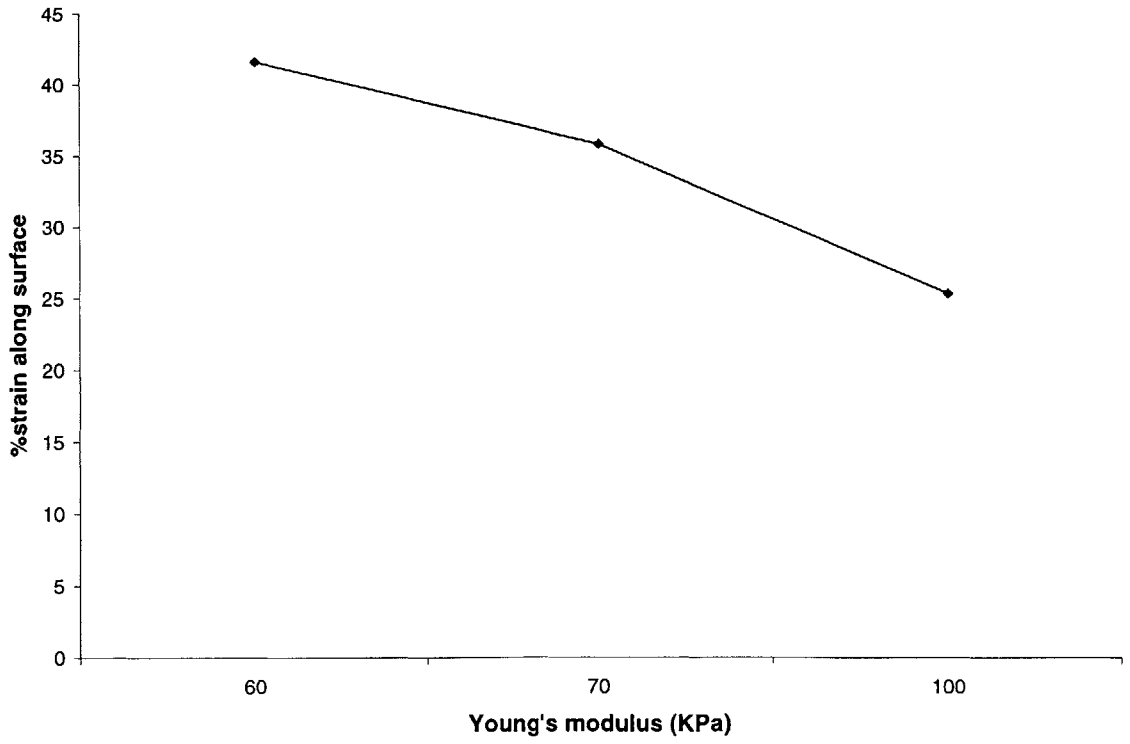
14) Meshing → mesh surfaces → nodes per element: 3, element group: 2 (porous), surface (porous) 1. Then add → repeat for element group 1 (non porous) and apply to surfaces 2 and 3.

15) Save as data file

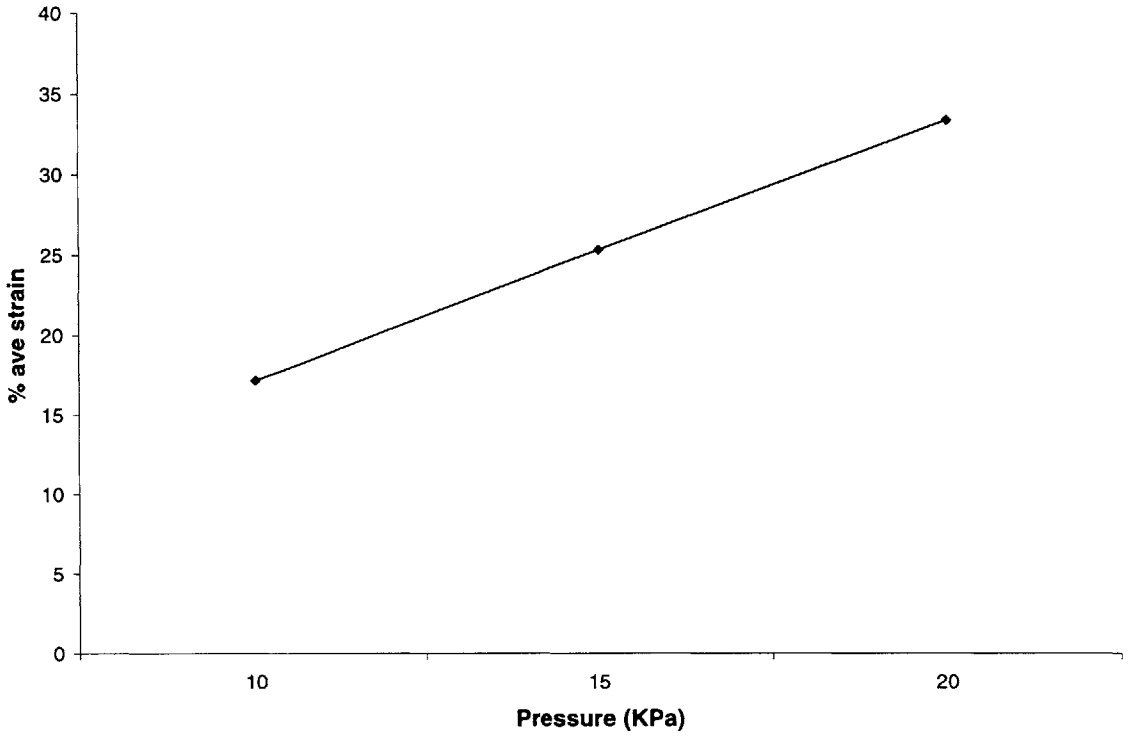
16) Run the FSI by going to Solution → Launch ADINA FSI → use ctrl to choose the two files → OK

Results

Average surface strain versus E

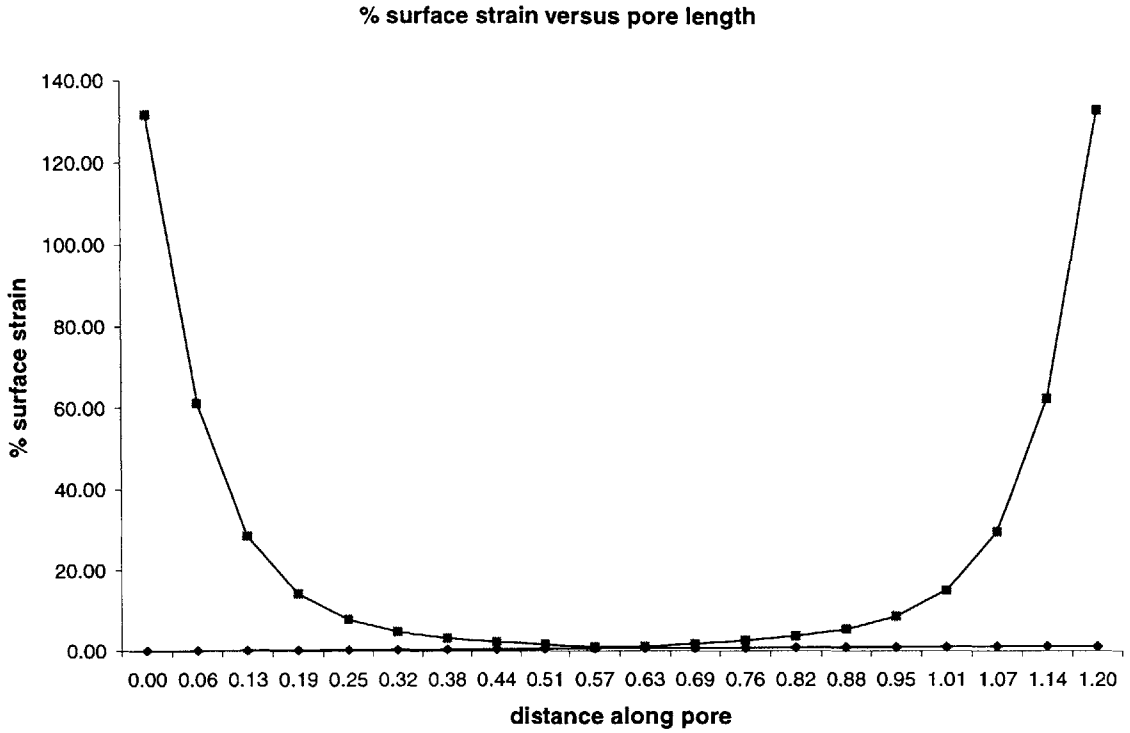


Average surface strain versus Pressure

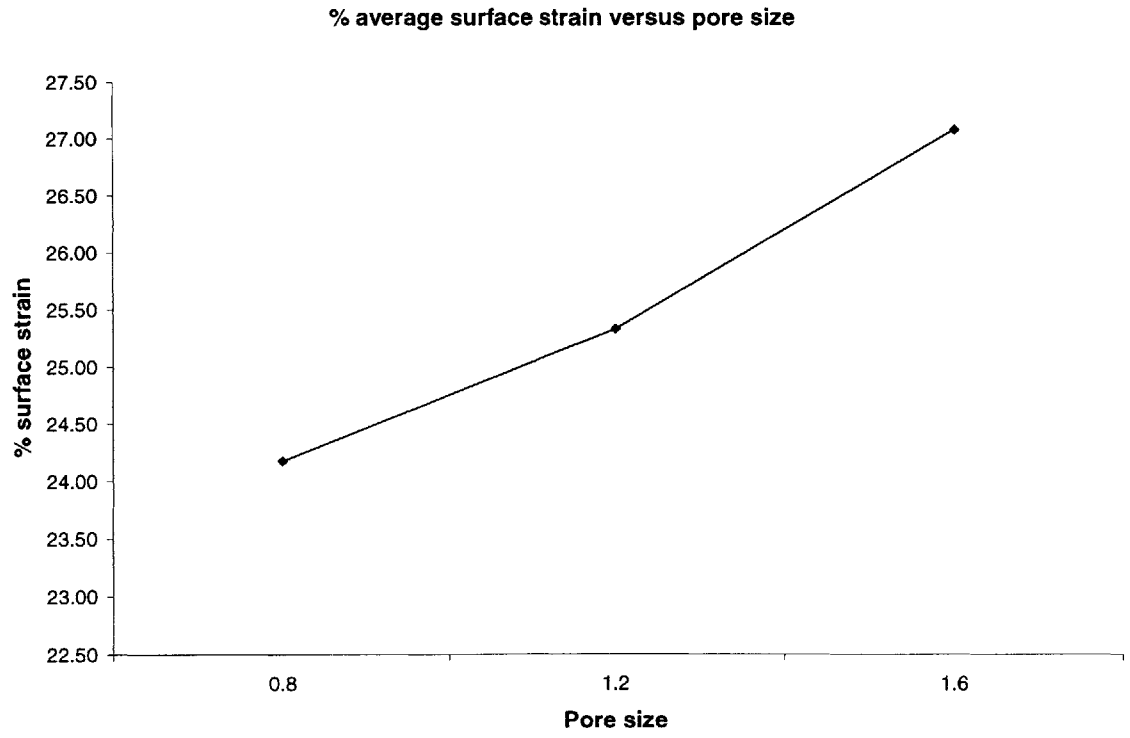


Viscosity and permeability were also changed, but they showed very little variation in average strains (however, they did show changes in velocity). Further, the model became unstable over values of Poisson's ratio that were lower than 0.47. Thus, the values of Poisson's ratio that were obtained were over too small a variation.

Next, let's look at variation of strain over the length of the pore

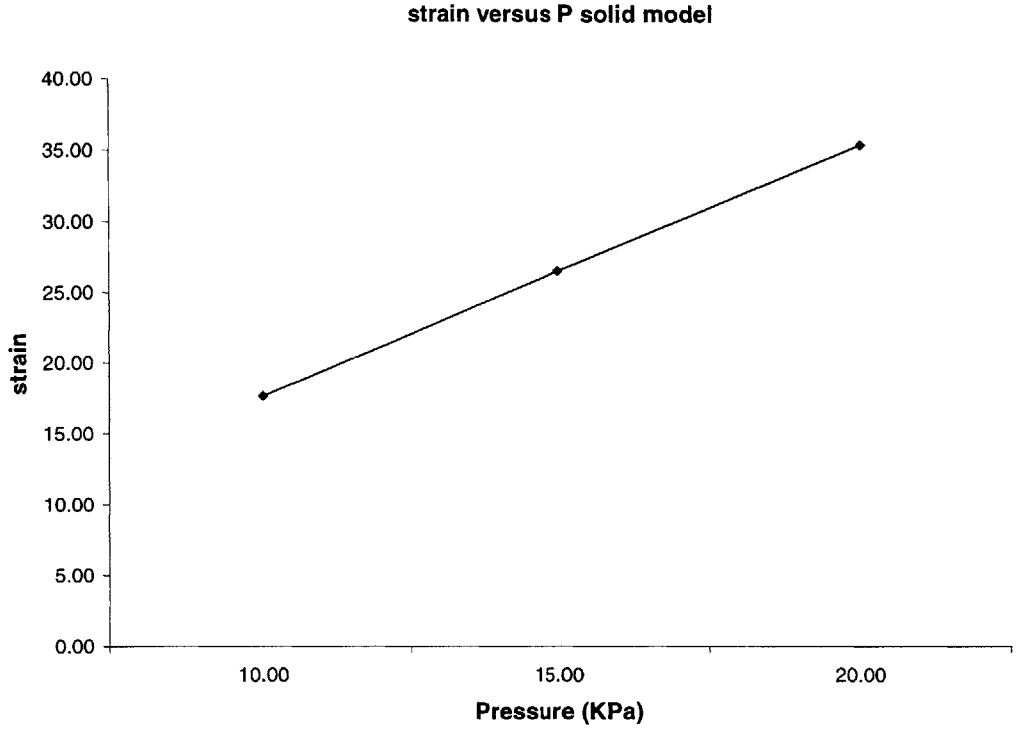


Next we look at the % surface strain versus pore size in the biphasic model

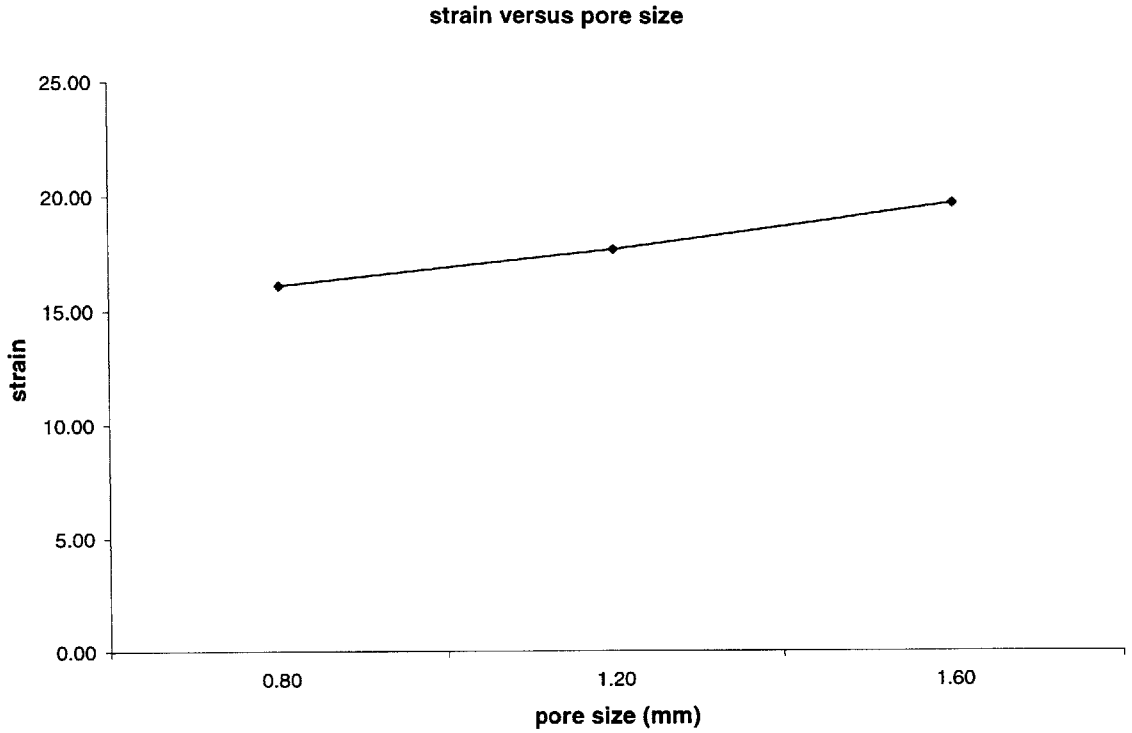


Next we look at some of these parameters for the solid model to see if we have similar strains or different stains (we obtain similar strains).

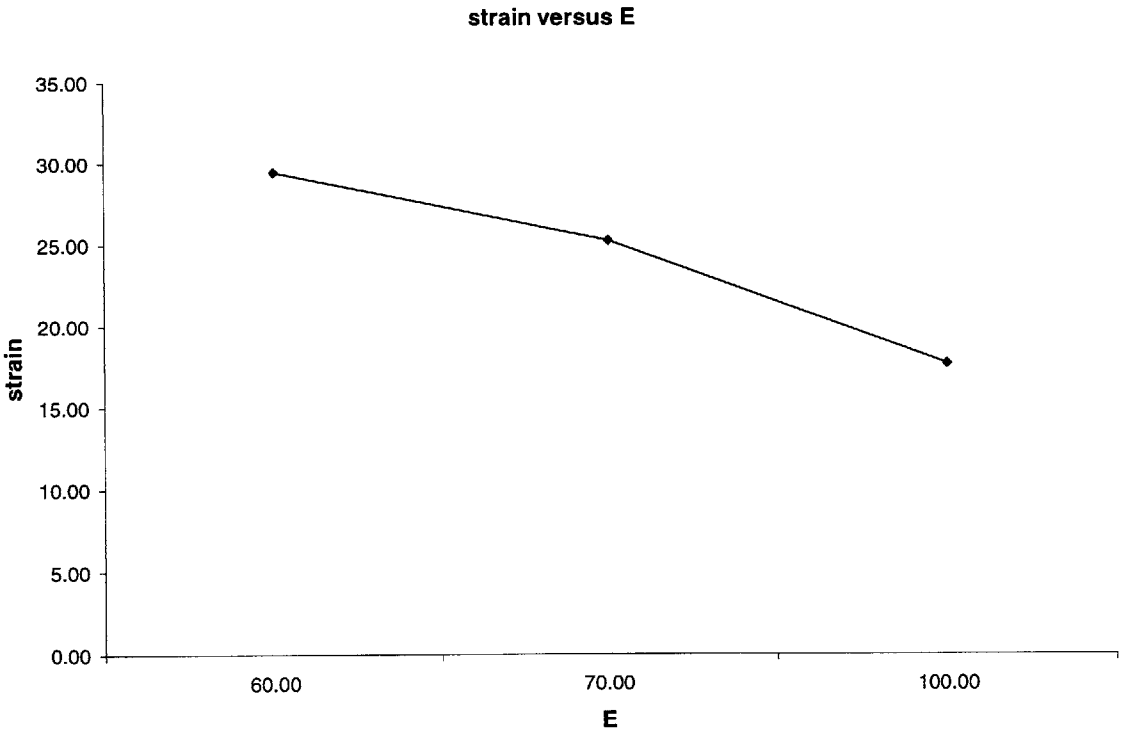
Surface strain versus pressure in solid model



Surface strain versus pore size in solid model



Strain versus E in the solid model



Conclusions

We see that the strains are very similar. Thus, the solid approximation is quite appropriate.

Glossary

Isotropic/anisotropic Assumes the stiffness of a material is the same along any axis. A biologicexample of an isotropic material is fat. In contrast, many biologic materials are anisotropic:they are very strong along one axis and weak perpendicular to that axis. For example, a tendonis very stiff along its longitudinal axis and much less stiff across its transverse axis.

Linear Assumes data can be fit to an equation of a line. For most biologic materials such as skin, the relationship between stress andstrain is nonlinear and exponential in nature over large strains. To simplify the model, a linear stress-strain relationship is assumed over small displacements.

Pore volume fraction (Vf) Proportion of sponge occupied by air; pore diameter divided by the pore diameter plus strut thickness.

Fold change Ratio of expression values (for example stretch to control).

Morphogenesis (cell biology) The process of shape formation: the processes that are responsible for producing the complex shapes of adults from the simple ball of cells that derives from division of the fertilized egg (From online medical dictionary).

Suture Closing a wound with suitable biocompatible 'thread'. In general, a suture can be applied to non biological tissue (as when we 'suture' the force-producing rubber band to the latex that is attached to the rat ear).

Biological replicate/technical replicate A biological replicate refers to multiple biological samples, while a technical replicate refers to multiple microarray runs.

Granulation: One of the small, red, grainlike prominences which form on a raw surface (that of wounds or ulcers), and are the efficient agents in the process of healing. The act or process of the formation of such prominences.

Keratinocyte: <pathology> skin cell, of the keratinized layer of epidermis: its characteristic intermediate filament protein is cytokeratin.

Focal adhesions: <cell biology> Areas of close apposition and thus presumably anchorage points, of the plasma membrane of a fibroblast (for example) to the substratum over which it is moving; Usually 1m x 0.2 m [sic—online medical dictionary] with the long axis parallel to the direction of movement, always associated with a cytoplasmic microfilament bundle that is attached via several proteins to the plasma membrane at an area of high protein concentration (this is noticeably electron dense in electron micrographs). Focal adhesions tend to be characteristic of slow moving cells.

Microfilament: <cell biology> Cytoplasmic filament of F actin (5-7nm) which function in structure and movement in eukaryotic cells. They may be laterally associated with other proteins (tropomyosin, alpha actinin) in some cases and may be anchored to the membrane. Microfilaments are conspicuous in adherens junctions.

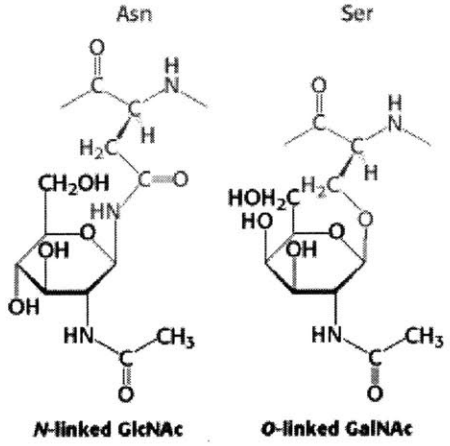
Microtubule: <cell biology> Cytoplasmic tubule, 25nm outside diameter with a 5nm thick wall. Made of tubulin heterodimers packed in a three start helix (or of 13 protofilaments looked at another way) and associated with various other proteins (MAPs, dynein, kinesin). Microtubules of the ciliary axoneme are more permanent than cytoplasmic and spindle microtubules. (online medical dictionary)

"Microtubules are components of all eukaryotic cells. They participate in a range of motion processes, such as:

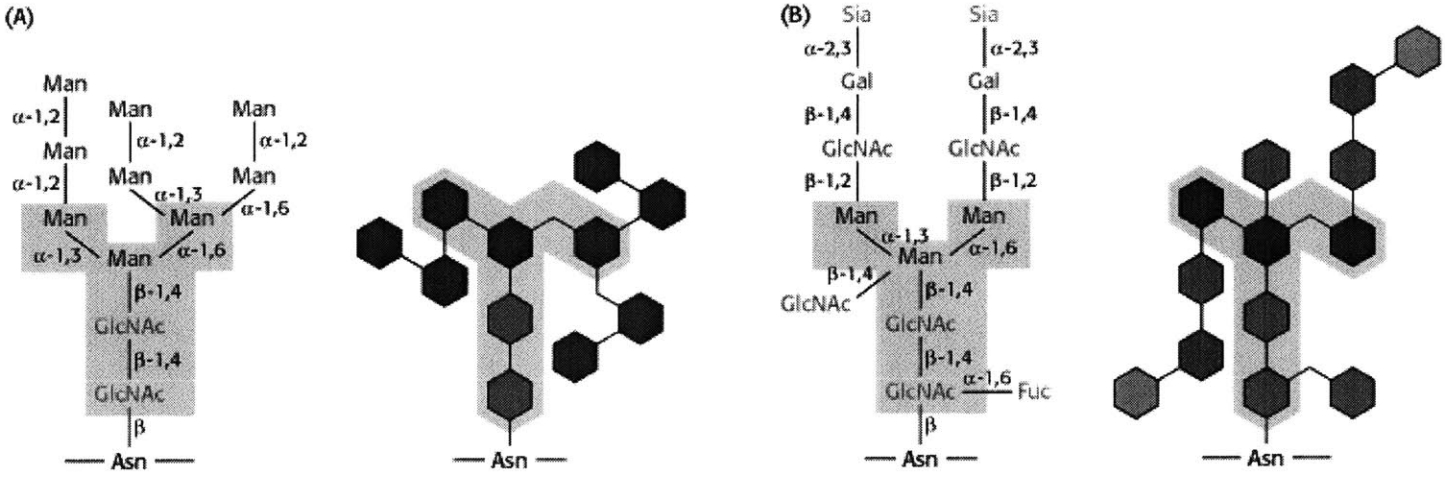
- the movements of flagella and cilia,
- the movements of chromosomes during meiosis and mitosis [being part of the spindle] and
- the transport of granules and vesicles within the cells that effects cell wall formation, shape and specialization of the cells" (15).

Paracrine: <endocrinology> Form of signalling in which the target cell is close to the signal releasing cell. Neurotransmitters and neurohormones are usually considered to fall into this category.

Glycoproteins: Conjugated proteins containing one or more covalently linked carbohydrate residues. While technically describing conjugates in which the carbohydrate is less than 4 per cent by weight, the term is often used generically to include the mucoproteins and proteoglycans (19).

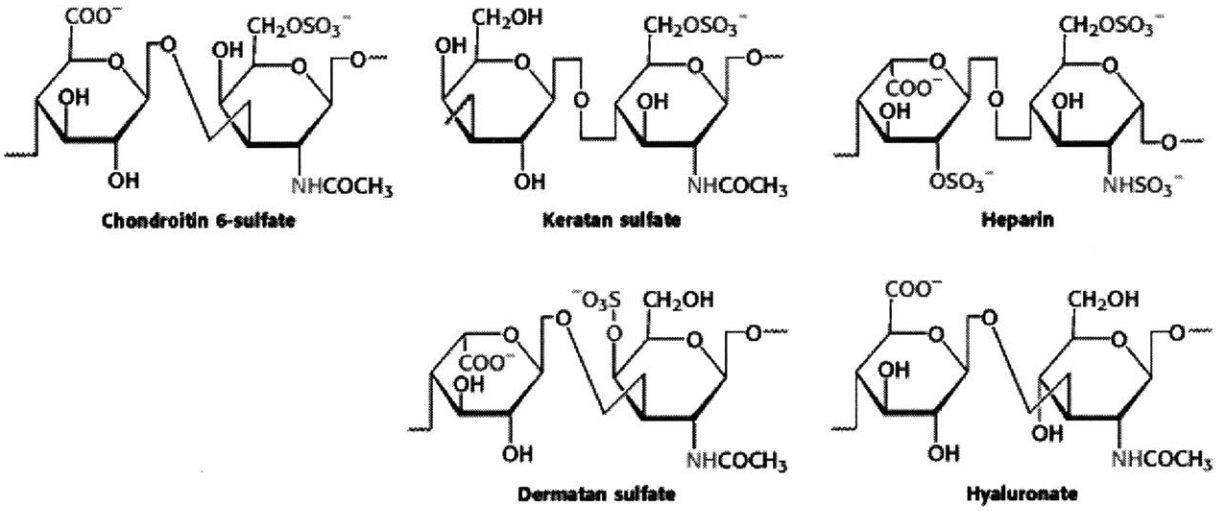


http://www.ncbi.nlm.nih.gov/books/bv.fcgi?call=bv.View..ShowSection&rid=stryer.figgrp.1533
Figure above and below: Carbohydrates Can Be Attached to Proteins to Form Glycoproteins. Two specific examples are given next



Proteoglycans: Glycoproteins which have a very high polysaccharide content (on-line medical dictionary—(19)).

Glycosaminoglycans (Stryer 5th ed—(18)): Glycosaminoglycans Are Anionic Polysaccharide Chains Made of Repeating Disaccharide Units (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Search&db=books&doptcmdl=GenBookHL&term=proteoglycans+AND+stryer%5Bbook%5D+AND+215807%5Buid%5D&rid=stryer.section.1517#1526) – Stryer 5th ed online.



http://www.ncbi.nlm.nih.gov/books/bv.fcgi?call=bv.View..ShowSection&rid=stryer.figgrp.1527

Figure: Repeating units in different glycosaminoglycans

The glycan of a proteoglycan is a glycosaminoglycan, but a glycosaminoglycan can exist separately from a protein. Thus, a GAG is either just a GAG or is part of a PG. However, in cartilage PGs can be linked to a GAG – hyaluronan (20).

Appendix

P value finder with T test sorting

```
--------------------------
#!/usr/bin/perl -w
use strict;

#this code computes the P value for a given list of genesets. Our final result from
#this code is a table of P values versus geneset names. We will first obtain ES for each
geneset
#in a particular permutation. This will be stored in a 2D array that has geneset in first
column
#and ES in second column. This list will then be sorted largest to smallest to give us the
geneset
#with maxES for that permutation. This geneset name will be pushed into an array called
count. We
#will finally count the values of each geneset in this array. The inverse of each of these
will
#be the P values for each of these genesets. We will write a subroutine within the 1000
for loop
#that will return just one scalar argument (a string) that will be the MES geneset in each
iteration
#of the permute.

#first the 1000 loop
my @data1;
my @new;
my @count;
my @n;
my $k;
```

```
my @new1;


for(my $asub=0; $asub<=999; $asub++){
print "loop $asub \n";
push (@count, &ESfinder);
print "\n \n \n";
print "$count[$asub] \n";
print "\n \n \n";

} # end of global permute

#load geneset filenames into data1 array
my $gsetfiles = "gsetfiles.txt";
open(FH,"<$gsetfiles") or die("can't open input file $gsetfiles.txt: $!\n");
while (my $line=<FH>){
chomp($line);
push(@data1, $line);
}
close FH;

#count the number of times each geneset occurs in array count.
#now we count
for($k=0;$k<=$#data1;$k++){
        $new[$k]=0;
        for ($a=0;$a<=999;$a++){ #here we've used 1000, since we're only getting one
max value per permute
                if($data1[$k] eq $count[$a]){
                $new[$k]=$new[$k]+1;
#               print "$new[$k] \n";
                }
```

```
        }
}


#finally output results to a file (the P value next to gene name)
for(my $a=0;$a<=$#data1;$a++){
        $new1[$a]=$new[$a]/1000;
#       print "$allES[$a][0] \t $n[$a] \n";
#       print "$allES[$a][0] \t $new[$a] \n";
}

#print max enrichment score
for (my $b=0;$b<=$#data1;$b++){
        print "geneset $data1[$b]  \t has P value  \t $new1[$b] \t & \t $new[$b] \n";
}

#check count if all there
# for(my $check=0;$check<=999;$check++){
#       print "$count[$check] \t $check \n";
# }


#print results to a file

open OUTPUTFILE, ">outputfile.txt";

for (my $a=0;$a<=$#data1;$a++){
print OUTPUTFILE "$data1[$a]  \t has P value  \t $new1[$a] \t & \t $new[$a] \n";
}

close OUTPUTFILE;
```

```
#-------------------------------------------------------------------
#-------------------------------------------------------------------
# now define the subroutine

sub ESfinder {
#load expression1.txt into a matrix with 3 columns

my @data;
my @parsed;
my $sum;
my @pES;
my @sort_pES;
my $mES;
my $s;
my $i;
my @psort_new;
my @data1;
my $gsetfiles;

#We'll have to modify the expression1 text file to have all the eight time points

my $expression1="expression1.txt";
open(EXPRESSION1,$expression1) or die("can't open input file $expression1: $!\n");
while(<EXPRESSION1>){
    chomp($_);
    push(@data, $_);
}
close EXPRESSION1;

for (my $i=0; $i<=$#data; $i++) {
```

```
    push(@parsed, [split(/\t/, $data[$i])])
}

print "number  in permute \n";

#save $#data in $datalength and kill @data
my $datalength=$#data;
@data=();

#our code will be different from here on down
#============================================================
==
#============================================================
==
#generate 8 random numbers corresponding to each column in the stretch system

#generate an array with first column being integers 9 through 16
my @StretchSorter;
my @Stretch_Sorted;
for (my $a=0;$a<=7; $a++){
        $StretchSorter[$a][0]=$a+9;
        $StretchSorter[$a][1]=rand();
}

# for (my $a=0;$a<=7; $a++){
# print "$StretchSorter[$a][0] \t $StretchSorter[$a][1] \n";
# }

print "\n \n";
#sort StretchSorter
@Stretch_Sorted=sort{$b->[1] <=> $a->[1]}@StretchSorter;
```

```
# for (my $a=0;$a<=7; $a++){
# print "$Stretch_Sorted[$a][0] \t $Stretch_Sorted[$a][1] \n";
# }

#-------------------------------------------------------------
#define the swap algorithm
#start with a random number for each column in the control set
#if number greater than 0.5 swap otherwise not swap
#this defines whether we swap or not

my @ControlTracker;
my @Tsquare;


for(my $a=0;$a<=7;$a++){
        my $random=rand();
        if ($random > 0.5){
                $ControlTracker[$a]=$Stretch_Sorted[$a][0];
                $Stretch_Sorted[$a][0]=$a+1;
        }
        else {$ControlTracker[$a]=$a+1;}
}

#permute the matrix
for(my $a=0;$a<=$datalength;$a++){
        my $Xsum=0;
        my $Xmean=0;
        my $Ysum=0;
        my $Ymean=0;
# the next loop computes the sum of each group
```

```
	for(my $b=0;$b<=7;$b++){
		$Xsum+=$parsed[$a][$ControlTracker[$b]];
		$Ysum+=$parsed[$a][$Stretch_Sorted[$b][0]];
	}

#now we compute the means
$Xmean=($Xsum/8);
$Ymean=($Ysum/8);

#calculate the t statistic denominator summation
##first make an intermediate variable of Ybar-Xbar
my $Sumhelper=$Ymean-$Xmean;
##
#now the summation
my $Denom_sum=0;
my $Denom_sum_square=0;
for(my $b=0;$b<=7;$b++){
	$Denom_sum=($Sumhelper+$parsed[$a][$ControlTracker[$b]]-
$parsed[$a][$Stretch_Sorted[$b][0]]);
	$Denom_sum_square+=($Denom_sum*$Denom_sum);
}

$Tsquare[$a][0]=$a;
$Tsquare[$a][1]=($Sumhelper*$Sumhelper)*(8*7)/($Denom_sum_square);

#close the a loop
}

@parsed=();
```

```
#=================================================================
===
#=================================================================
===

#sort
@psort_new=sort{$b->[1] <=> $a->[1]}@Tsquare;
print "array sorted \n";

# for (my $newloop=0; $newloop<=$datalength; $newloop++){
#       print "$parsed[$newloop][3] \t $parsed[$newloop][2] \t $parsed[$newloop][1]
\n";
#       }

#next do a for loop from 1 to number of genesets. This can be found by loading the file
called
#gsetfiles.txt. This will load all the genesets into array data1. Next we will load each
geneset file
#from data1 into vector called geneset and find the running sum. A sorting of this will
give ES.
#This ES will be 'pushed' into a 2D array with first column gset name and the second the
ES. At
#the end of the for loop the array geneset will be set to () and then undef.

#so first load genesetfile into array data1.
$gsetfiles = "gsetfiles.txt";
open(FH,"<$gsetfiles") or die("can't open input file $gsetfiles.txt: $!\n");
while (my $line=<FH>){
chomp($line);
push(@data1, $line);
}
```

```
close FH;
print "genesetfile loaded \n";

#now loop from 0 to number of genesets. In each loop load the i'th file and then compute
the
#running sum and then sort this file to obtain the ES for this geneset. push this number
into a
#2D array with the first column being genesetname and the second the ES
my @sumArray;
my @sumArraySort;
for($i=0; $i<=$#data1; $i++){
        my @gsetarray;
        print "$i \n";
        open(FH1, "<$data1[$i]");
        while(<FH1>){
                chomp($_);
                push(@gsetarray,$_);
         }
#remember now that our gsetfiles are all data entries long
        #now find the running sum of @gsetarray by referencing the psort_new array
indices
        for($s=0;$s<=$#gsetarray;$s++){
#               print "$s \n";
                if ($s==0){
                        $sum=$gsetarray[$psort_new[0][0]];
                }
                else {
                        $sum=$gsetarray[$psort_new[$s][0]]+$sum;
                }
                $sumArray[$s]=$sum;
        }
```

```
        @sumArraySort = sort { $b <=> $a } @sumArray;
        $pES[$i][0] = $data1[$i];
        $pES[$i][1] = $sumArraySort[0];
        $sum=0;
        @gsetarray=();
        @sumArray=();
        @sumArraySort=();
}

print "sum computed \n";
#now we sort @pES by the column number 1 (with ES scores). We pull off the first entry
in
#column 0. This is the gsetname that is sent back from the subroutine.
@sort_pES=sort{$b->[1] <=> $a->[1]}@pES;
$mES=$sort_pES[0][0];
#print "$mES \n";
#print "$sort_pES[0][0] \n";
#print "$pES[0][0] \n";
print "$mES \n";
return $mES;
} # end of subroutine



# #good way to check
# #if ($a % 5000 == 0) { print "$a...\n";}
# if (!defined($value)) { print "$value is undefined";}
# if it's a show stopper, you can use die("error message"); instead

-------------------------------------------------------
```

Appendix

Creating genesets correctly formatted (see text in thesis for what this means)

The following code will create different gene set files with the Kolmogorov statistic on each line corresponding to its gene (see main body of thesis for explanation)

```
#!/usr/bin/perl

#this program requires the following input files:
#gsetfiles.txt (this file has list of all gsetfiles--eg. gset1.txt, gset2.txt etc one on each line
#the gset.txt files should only have gene names not squared values
#expression.txt (just one time point expression set)

$expression="expression.txt";
open(EXPRESSION,$expression) or die("can't open input file $expression: $!\n");
while(<EXPRESSION>){
    chomp($_);
    push(@data, $_);
}
close EXPRESSION;

for (my $i=0; $i<=$#data; $i++) {
    push(@parsed, [split(/\t/, $data[$i])])
 }

 print "LOOPS 0000000000 \n";

$gsetfiles = "gsetfiles.txt";
```

```
open(FH,"<$gsetfiles") or die("can't open input file $gsetfiles.txt: $!\n");
while (my $line=<FH>){
chomp($line);
push(@data1, $line);
}
close FH;
for($i=0; $i<=$#data1; $i++){
        open(FH1, "<$data1[$i]");
        while(<FH1>){
                chomp($_);
                push(@gsetarray,$_);
         }
        foreach $j (@gsetarray) {
           push(@tempgsetarray, $j) unless ($seen{$j}++);
        }
        @gsetarray = @tempgsetarray;
        @tempgsetarray=();
        %seen=();
   $length[$i]=$#gsetarray;
        for($b=0;$b<=$#gsetarray;$b++){
                $array[$i][$b+1]=$gsetarray[$b];
        }
        @gsetarray=();
}

for($a=0;$a<=$#data;$a++){
        $ExampleGlobal[$a]=$parsed[$a][0];
}

for ($a=0;$a<=$#data1;$a++){
        for($b=1;$b<=$length[$a]+1;$b++){
```

```
                $ExampleSubset[$b-1]=$array[$a][$b];
        }

                my %ExampleGlobalHash;
                foreach my $i (0..$#ExampleGlobal) {
                if (!defined($ExampleGlobalHash{$ExampleGlobal[$i]})) {
        $ExampleGlobalHash{$ExampleGlobal[$i]} = $i;
                }
                }
                foreach my $j (0..$#ExampleSubset) {
                if (defined($ExampleGlobalHash{$ExampleSubset[$j]})) {
                        $index=$ExampleGlobalHash{$ExampleSubset[$j]};
                $parsed[$index][$a+4]=1;
        }
        }
        @ExampleSubset=();
}

print "loop 22222222222 \n";

#------------------------end of geneset loading----------
#------------add zeroes to the rest of the matrix---------
for($a=0;$a<=$#data;$a++){
        for($b=4;$b<=$#data1+4;$b++){
                if($parsed[$a][$b]!=1){
                        $parsed[$a][$b]=0
                }
        }
}
#---------the above is an exercise in futility but makes the code run with the previous code
```

```
#create the difference column
for($c=0; $c<=$#data; $c++){
        $parsed[$c][3]=$parsed[$c][2]-$parsed[$c][1];
}

#find the statistics

print "LOOPS 3333333333 \n";

for($aaaaa=0;$aaaaa<=$#data1;$aaaaa++){
                for($bbbbb=0;$bbbbb<=$#data;$bbbbb++){
                        $XNG=-sqrt(($length[$aaaaa]+1)/($#data-$length[$aaaaa]));
                        $XG=sqrt(($#data-$length[$aaaaa])/($length[$aaaaa]+1));
                        if ($parsed[$bbbbb][$aaaaa+4] == 1){
                        $parsed[$bbbbb][4+$aaaaa]=$XG;
                        }
                        else {
                        $parsed[$bbbbb][4+$aaaaa]=$XNG;
                        }
                }
}

#print gset entries to different gset files
#we must remember that @data1 has the list of all genenames indexed from 0 to $#data1
#-------------------------------------------------------------------------------

for($i=0; $i<=$#data1; $i++){
        open(FH2, ">$data1[$i]");
        for ($a=0;$a<=$#data;$a++){
                if ($a != 0){print FH2 "\n";}
```

```
            print FH2 "$parsed[$a][4+$i]";
        }
        close FH2;
}
```

Appendix

Combined up and down regulated ES scores and P values for 3 iterations

Gene Set	Enrichment Score	P value 1st iter.	P value 2nd iter.	P value 3rd iter.
hypoxia.txt	321.0	0.001	0.003	0.007
response_to_mechanical_stimulus.txt	310.3	0.014	0.015	0.016
c3_U133_probes.txt	308.1	0.018	0.018	0.018
MAP00340_Histidine_metabolism.txt	304.6	0.003	0.004	0.004
c27_U133_probes.txt	301.7	0.016	0.016	0.017
MAP00280_Valine_leucine_and_isoleucine_degradation.txt	283.6	0.008	0.011	0.01
c23_U133_probes.txt	278.6	0.013	0.01	0.012
c6_U133_probes.txt	254.7	0.01	0.017	0.014
mitochondr_HG-U133A_probes.txt	252.7	0.005	0.01	0.009
MAP00140_C21_Steroid_hormone_metabolism.txt	240.8	0.002	0.006	0
MAP00740_Riboflavin_metabolism.txt	238.7	0.006	0.008	0.004
Hum_Fb_Serum_EarlyTF.txt	237.6	0.008	0.002	0.007
MAP00562_Inositol_phosphate_metabolism.txt	228.1	0.014	0.006	0.011
MAP00480_Glutathione_metabolism.txt	227.3	0.002	0.002	0.005
c12_U133_probes.txt	220.3	0.008	0.008	0.006
MAP00010_Glycolysis_Gluconeogenesis.txt	219.5	0.006	0.014	0.011
cluster1_LPS_mouse_urinary.txt	217.8	0.007	0.006	0.008
ROS_HG-U133A_probes.txt	217.1	0.007	0.004	0.006
MAP03030_DNA_polymerase.txt	210.3	0.002	0.006	0.001
c9_U133_probes.txt	209.7	0.014	0.018	0.014
MAP00380_Tryptophan_metabolism.txt	209.1	0.003	0	0.001
MAP00380_Tryptophan_metabolism~.txt	209.1	0	0	0
c20_U133_probes.txt	205.8	0.004	0.003	0.003
MAP00632_Benzoate_degradation.txt	202.7	0.007	0.006	0.003
c34_U133_probes.txt	196.4	0.006	0.005	0.013
MAP00790_Folate_biosynthesis.txt	194.0	0.001	0.002	0.004
Hum_Fb_Serum_Coag_Hemo.txt	190.9	0.005	0.006	0.007
inflammatory.txt	188.7	0.006	0.008	0.009
c15_U133_probes.txt	187.6	0.013	0.007	0.015
human_mitoDB_6_2002_HG-U133A_probes.txt	187.2	0.018	0.015	0.016
MAP00625_Tetrachloroethene_degradation.txt	186.4	0	0	0
c8_U133_probes.txt	185.7	0.012	0.007	0.009
MAP00521_Streptomycin_biosynthesis.txt	182.8	0.001	0	0
c10_U133_probes.txt	180.4	0.002	0.007	0.011
MAP00720_Reductive_carboxylate_cycle_CO2_fixation.txt	179.2	0.003	0.003	0.001
c1_U133_probes.txt	179.1	0.016	0.011	0.008
c30_U133_probes.txt	177.9	0.008	0.007	0.004
MAP00310_Lysine_degradation.txt	175.1	0.001	0.003	0.002
c16_U133_probes.txt	175.0	0.008	0.006	0.004

c26_U133_probes.txt	171.4	0.008	0.005	0.006
MAP00830_Retinol_metabolism.txt	169.3	0	0	0
GO_0005739_HG-U133A_probes.txt	168.9	0.001	0.004	0.003
c19_U133_probes.txt	167.6	0.011	0.017	0.009
MAP00031_Inositol_metabolism.txt	166.5	0	0	0
c17_U133_probes.txt	166.5	0.003	0	0
MAP00530_Aminosugars_metabolism.txt	166.0	0.001	0	0.002
c5_U133_probes.txt	165.5	0.013	0.008	0.014
MAP00440_Aminophosphonate_metabolism.txt	165.4	0	0	0
MAP00460_Cyanoamino_acid_metabolism.txt	165.4	0	0	0
MAP00360_Phenylalanine_metabolism.txt	164.8	0.003	0.006	0.005
ECM_constituent_conferring_elasticity.txt	164.8	0	0	0
ECM_structural_constituent_conferring_compression_resistance.txt	164.8	0	0	0
INS_HG-U133A_probes.txt	163.5	0.009	0.007	0.009
MAP00020_Citrate_cycle_TCA_cycle.txt	161.7	0.001	0.002	0.003
c32_U133_probes.txt	159.7	0.008	0.001	0.003
MAP00522_Erythromycin_biosynthesis.txt	159.1	0.001	0	0.001
Hum_Fb_Serum_Chol_Biosyn.txt	158.7	0.001	0	0.001
c24_U133_probes.txt	156.6	0.014	0.013	0.014
cluster3_LPS_mouse_urinary.txt	154.1	0.002	0.002	0.006
c7_U133_probes.txt	154.1	0.068	0.069	0.065
MAP00361_gamma_Hexachlorocyclohexane_degradation.txt	153.1	0.003	0.005	0.003
GLUCO_HG-U133A_probes.txt	151.5	0	0.002	0.002
GLYCOL_HG-U133A_probes.txt	149.3	0	0.004	0.003
cluster13_LPS_mouse_urinary.txt	147.5	0.002	0.006	0.003
MAP00220_Urea_cycle_and_metabolism_of_amino_groups.txt	147.1	0.006	0.001	0.002
MAP00630_Glyoxylate_and_dicarboxylate_metabolism.txt	147.0	0.002	0.001	0.001
cluster12_LPS_mouse_urinary.txt	146.7	0.004	0.003	0.006
MAP00071_Fatty_acid_metabolism.txt	143.5	0.011	0.014	0.015
MAP00511_N_Glycan_degradation.txt	141.7	0.001	0.003	0.005
MAP00150_Androgen_and_estrogen_metabolism.txt	140.8	0.006	0.004	0.007
MAP00251_Glutamate_metabolism.txt	140.4	0.002	0.004	0.001
MAP00072_Synthesis_and_degradation_of_ketone_bodies.txt	139.1	0	0.002	0
KET_HG-U133A_probes.txt	139.1	0.006	0.003	0.005
MAP00051_Fructose_and_mannose_metabolism.txt	135.9	0	0.003	0.001
MAP00053_Ascorbate_and_aldarate_metabolism.txt	135.8	0	0	0
MAP00631_1_2_Dichloroethane_degradation.txt	135.8	0	0	0
c18_U133_probes.txt	134.9	0.032	0.034	0.023
c0_U133_probes.txt	134.8	0.014	0.012	0.025
MAP00512_O_Glycans_biosynthesis.txt	133.9	0	0.001	0.001
c21_U133_probes.txt	133.5	0.002	0	0.001
cluster7_LPS_mouse_urinary.txt	132.9	0.006	0.005	0.004
MAP00410_beta_Alanine_metabolism.txt	130.4	0.002	0.002	0.001

cluster9_LPS_mouse_urinary.txt	128.6	0.006	0.006	0.005
MAP00770_Pantothenate_and_CoA_biosynthesis.txt	128.6	0	0	0
c25_U133_probes.txt	127.0	0.008	0.012	0.006
MAP00940_Flavonoids_stilbene_and_lignin_biosynthesis.txt	126.1	0	0.001	0.001
MAP00760_Nicotinate_and_nicotinamide_metabolism.txt	125.9	0.003	0.005	0.003
c2_U133_probes.txt	123.0	0.062	0.063	0.063
GLYCOGEN_HG-133A_probes.txt	122.2	0.002	0.002	0.001
MAP00130_Ubiquinone_biosynthesis.txt	119.6	0.001	0	0.002
PYR_HG-U133A_probes.txt	117.0	0.003	0	0.005
MAP00640_Propanoate_metabolism.txt	116.2	0.001	0.004	0.001
FA_HG-U133A_probes.txt	114.3	0.008	0.005	0.004
MAP00030_Pentose_phosphate_pathway.txt	112.2	0.003	0.003	0.003
c22_U133_probes.txt	111.5	0.002	0.003	0.004
MAP00600_Sphingoglycolipid_metabolism.txt	110.5	0.007	0.004	0.006
c33_U133_probes.txt	108.2	0.007	0.011	0.012
c31_U133_probes.txt	105.2	0.009	0.008	0.005
MAP00750_Vitamin_B6_metabolism.txt	103.6	0.003	0.002	0.002
MAP00620_Pyruvate_metabolism.txt	103.6	0.002	0.002	0.002
MAP00710_Carbon_fixation.txt	103.3	0.005	0.013	0.006
cluster4_LPS_mouse_urinary.txt	100.9	0.006	0.004	0
MAP00510_N_Glycans_biosynthesis.txt	99.4	0.006	0.002	0.004
c35_U133_probes.txt	98.5	0.007	0.01	0.008
MAP00970_Aminoacyl_tRNA_biosynthesis.txt	98.4	0.01	0.016	0.014
MAP00062_Fatty_acid_biosynthesis_path_2.txt	98.0	0.002	0.002	0.001
inflammatory_mechanical.txt	97.7	0.008	0.004	0.018
MAP00052_Galactose_metabolism.txt	97.7	0.002	0	0.001
MAP00471_D_Glutamine_and_D_glutamate_metabolism.txt	97.0	0.001	0.002	0.001
c11_U133_probes.txt	96.6	0.002	0.005	0.006
OXPHOS_HG-U133A_probes.txt	96.4	0.033	0.039	0.027
MAP00570_Sphingophospholipid_biosynthesis.txt	93.8	0	0	0
MAP00260_Glycine_serine_and_threonine_metabolism.txt	92.4	0.003	0.001	0.002
MAP00960_Alkaloid_biosynthesis_II.txt	89.1	0.004	0.002	0.001
MAP00643_Styrene_degradation.txt	88.0	0.001	0.001	0.002
MAP00350_Tyrosine_metabolism.txt	87.9	0.004	0.004	0.003
MAP00533_Keratan_sulfate_biosynthesis.txt	86.7	0.002	0.003	0.006
perception_of_pain_sensory_transductuon_of_mechanical_stimulus.txt	85.7	0.002	0.002	0.005
sensory_perception_of_mechanical_stimulus.txt	85.7	0	0	0
cluster2_LPS_mouse_urinary.txt	85.5	0.006	0.004	0.002
MAP00252_Alanine_and_aspartate_metabolism.txt	85.1	0.003	0.001	0.005
MAP00330_Arginine_and_proline_metabolism.txt	83.2	0.004	0.007	0.006
MAP00580_Phospholipid_degradation.txt	82.8	0.008	0.009	0.004
MAP00650_Butanoate_metabolism.txt	81.0	0.003	0.004	0.006
MAP00450_Selenoamino_acid_metabolism.txt	80.5	0.001	0	0.001

c13_U133_probes.txt	79.5	0.002	0.005	0.004
MAP00561_Glycerolipid_metabolism.txt	79.4	0.012	0.002	0.004
MAP03020_RNA_polymerase.txt	79.4	0.001	0.003	0.004
MAP00240_Pyrimidine_metabolism.txt	79.3	0.005	0.003	0.007
MAP00300_Lysine_biosynthesis.txt	77.2	0	0	0.001
MAP00230_Purine_metabolism.txt	76.5	0.006	0.005	0.006
MAP00670_One_carbon_pool_by_folate.txt	74.5	0.001	0.001	0
c14_U133_probes.txt	74.4	0.012	0.004	0.005
MAP00120_Bile_acid_biosynthesis.txt	72.3	0.002	0.001	0.004
cluster6_LPS_mouse_urinary.txt	72.1	0.001	0.001	0.002
cluster11_LPS_mouse_urinary.txt	71.8	0.006	0.008	0.009
cluster5_LPS_mouse_urinary.txt	71.5	0.01	0.006	0.007
MAP00400_Phenylalanine_tyrosine_and_tryptophan_biosynthesis.txt	71.2	0.009	0.005	0.008
MAP00601_Blood_group_glycolipid_biosynthesis_lact_series.txt	68.8	0.005	0.004	0.003
MAP00272_Cysteine_metabolism.txt	64.2	0.001	0.006	0
MAP00860_Porphyrin_and_chlorophyll_metabolism.txt	63.8	0.001	0.003	0.008
MAP00472_D_Arginine_and_D_ornithine_metabolism.txt	59.2	0	0	0
MAP00100_Sterol_biosynthesis.txt	56.3	0.022	0.014	0.024
MAP00532_Chondroitin_Heparan_sulfate_biosynthesis.txt	52.5	0.001	0.003	0.007
MAP00271_Methionine_metabolism.txt	51.9	0	0.003	0.002
c29_U133_probes.txt	51.9	0.008	0.005	0.006
MAP00430_Taurine_and_hypotaurine_metabolism.txt	51.3	0.007	0.005	0.005
MAP00680_Methane_metabolism.txt	50.2	0.001	0.003	0.001
MAP00602_Blood_group_glycolipid_biosynthesis_neolact_series.txt	49.9	0.006	0.008	0.005
MAP00910_Nitrogen_metabolism.txt	46.5	0.004	0.006	0.003
MAP00061_Fatty_acid_biosynthesis_path_1.txt	41.1	0	0	0
MAP00253_Tetracycline_biosynthesis.txt	41.1	0	0	0
MAP00520_Nucleotide_sugars_metabolism.txt	36.9	0	0	0
MAP00195_Photosynthesis.txt	35.2	0.005	0.004	0.001
MAP00190_Oxidative_phosphorylation.txt	34.2	0.01	0.006	0.005
MAP00193_ATP_synthesis.txt	34.1	0.006	0.004	0.004
MAP03070_Type_III_secretion_system.txt	34.1	0	0	0
MAP00603_Globoside_metabolism.txt	33.9	0.006	0.007	0.002
MAP00040_Pentose_and_glucuronate_interconversions.txt	33.0	0.004	0.006	0.005
TCA_HG-U133A_probes.txt	32.9	0.014	0.008	0.003
cluster14_LPS_mouse_urinary.txt	32.1	0.002	0.002	0.006
MAP00531_Glycosaminoglycan_degradation.txt	27.3	0.006	0.002	0.001
cluster15_LPS_mouse_urinary.txt	26.8	0.009	0.006	0.006
MAP00290_Valine_leucine_and_isoleucine_biosynthesis.txt	25.9	0.001	0	0
MAP00500_Starch_and_sucrose_metabolism.txt	25.7	0	0.001	0
c4_U133_probes.txt	21.8	0.011	0.013	0.007
c28_U133_probes.txt	21.0	0.001	0.004	0.006

cluster10_LPS_mouse_urinary.txt	19.1	0.007	0.01	0.012
MAP00590_Prostaglandin_and_leukotriene_metabolism.txt	16.8	0.002	0.011	0.001
MAP00780_Biotin_metabolism.txt	12.1	0	0.001	0.002
MAP00920_Sulfur_metabolism.txt	9.2	0.001	0.002	0.002
MAP00900_Terpenoid_biosynthesis.txt	9.0	0.019	0.016	0.016
MAP00950_Alkaloid_biosynthesis_I.txt	5.6	0.002	0.004	0.002

Appendix

Up regulated ES with P values

Gene Set	Enrichment Score	P value
response_to_mechanical_stimulus.txt	560.2	0.085
c26_U133_probes.txt	415.6	0.014
c6_U133_probes.txt	343.6	0.023
c3_U133_probes.txt	336.8	0.003
Hum_Fb_Serum_EarlyTF.txt	335.1	0
OXPHOS_HG-U133A_probes.txt	328.9	0.131
c27_U133_probes.txt	289.0	0
mitochondr_HG-U133A_probes.txt	283.7	0.018
MAP00710_Carbon_fixation.txt	274.4	0.004
c14_U133_probes.txt	265.8	0.002
c30_U133_probes.txt	262.2	0.001
c18_U133_probes.txt	259.0	0.045
c16_U133_probes.txt	253.7	0.002
MAP00950_Alkaloid_biosynthesis_I.txt	249.4	0
MAP03020_RNA_polymerase.txt	249.1	0
c5_U133_probes.txt	244.7	0
c33_U133_probes.txt	244.3	0
c20_U133_probes.txt	239.7	0.003
GLYCOL_HG-U133A_probes.txt	223.9	0
ROS_HG-U133A_probes.txt	218.6	0
MAP00480_Glutathione_metabolism.txt	216.9	0.001
MAP00430_Taurine_and_hypotaurine_metabolism.txt	211.9	0.006
MAP00230_Purine_metabolism.txt	210.6	0.001
c10_U133_probes.txt	208.0	0.006
c12_U133_probes.txt	205.4	0.002
MAP00361_gamma_Hexachlorocyclohexane_degradation.txt	204.4	0.005
MAP00251_Glutamate_metabolism.txt	202.6	0.001
MAP00740_Riboflavin_metabolism.txt	202.1	0
c32_U133_probes.txt	202.0	0.002
c17_U133_probes.txt	200.6	0.001
c19_U133_probes.txt	197.8	0.004
MAP00940_Flavonoids_stilbene_and_lignin_biosynthesis.txt	192.8	0
MAP00970_Aminoacyl_tRNA_biosynthesis.txt	185.4	0.009
MAP00030_Pentose_phosphate_pathway.txt	184.4	0.002
MAP00010_Glycolysis_Gluconeogenesis.txt	183.3	0.001
MAP00190_Oxidative_phosphorylation.txt	182.9	0
c8_U133_probes.txt	181.8	0.019
MAP00031_Inositol_metabolism.txt	172.0	0
MAP00440_Aminophosphonate_metabolism.txt	171.5	0
MAP00460_Cyanoamino_acid_metabolism.txt	171.5	0
ECM_constituent_conferring_elasticity.txt	171.3	0
ECM_structural_constituent_conferring_compression_resistance.txt	171.3	0
c11_U133_probes.txt	169.0	0.001
c29_U133_probes.txt	168.1	0

c15_U133_probes.txt	166.7	0
c9_U133_probes.txt	166.5	0.004
GLUCO_HG-U133A_probes.txt	165.2	0.001
cluster10_LPS_mouse_urinary.txt	164.8	0
MAP00272_Cysteine_metabolism.txt	162.0	0
cluster1_LPS_mouse_urinary.txt	161.5	0.003
MAP00240_Pyrimidine_metabolism.txt	160.8	0.003
MAP00900_Terpenoid_biosynthesis.txt	160.3	0.007
c35_U133_probes.txt	154.2	0.001
MAP00130_Ubiquinone_biosynthesis.txt	152.1	0.001
cluster11_LPS_mouse_urinary.txt	151.6	0.009
MAP00040_Pentose_and_glucuronate_interconversions.txt	150.9	0
MAP00195_Photosynthesis.txt	150.6	0.001
c28_U133_probes.txt	145.2	0
MAP00720_Reductive_carboxylate_cycle_CO2_fixation.txt	143.3	0
MAP00400_Phenylalanine_tyrosine_and_tryptophan_biosynthesis.txt	141.9	0
MAP00193_ATP_synthesis.txt	140.1	0.001
MAP03070_Type_III_secretion_system.txt	140.1	0
MAP00570_Sphingophospholipid_biosynthesis.txt	138.6	0
TCA_HG-U133A_probes.txt	138.0	0.005
MAP00561_Glycerolipid_metabolism.txt	135.6	0
Hum_Fb_Serum_Coag_Hemo.txt	130.9	0.002
c21_U133_probes.txt	130.8	0.001
MAP00590_Prostaglandin_and_leukotriene_metabolism.txt	128.0	0
MAP00780_Biotin_metabolism.txt	125.6	0
cluster14_LPS_mouse_urinary.txt	125.4	0
PYR_HG-U133A_probes.txt	124.8	0
MAP00603_Globoside_metabolism.txt	122.7	0.002
MAP00472_D_Arginine_and_D_ornithine_metabolism.txt	122.3	0
MAP00100_Sterol_biosynthesis.txt	120.5	0.038
c7_U133_probes.txt	119.1	0.098
MAP00260_Glycine_serine_and_threonine_metabolism.txt	116.8	0.002
c31_U133_probes.txt	114.6	0.001
MAP03030_DNA_polymerase.txt	112.9	0
MAP00630_Glyoxylate_and_dicarboxylate_metabolism.txt	111.9	0
c34_U133_probes.txt	111.6	0.003
inflammatory_mechanical.txt	111.1	0
MAP00532_Chondroitin_Heparan_sulfate_biosynthesis.txt	111.0	0
MAP00625_Tetrachloroethene_degradation.txt	110.6	0
c0_U133_probes.txt	110.3	0.026
MAP00300_Lysine_biosynthesis.txt	109.5	0
cluster5_LPS_mouse_urinary.txt	109.5	0
MAP00790_Folate_biosynthesis.txt	108.1	0
c1_U133_probes.txt	106.1	0.062
MAP00920_Sulfur_metabolism.txt	105.5	0.001
c13_U133_probes.txt	105.2	0
MAP00471_D_Glutamine_and_D_glutamate_metabolism.txt	99.7	0
MAP00150_Androgen_and_estrogen_metabolism.txt	96.7	0
MAP00051_Fructose_and_mannose_metabolism.txt	94.8	0.003
cluster6_LPS_mouse_urinary.txt	94.0	0

MAP00360_Phenylalanine_metabolism.txt	89.2	0
MAP00860_Porphyrin_and_chlorophyll_metabolism.txt	89.0	0
MAP00522_Erythromycin_biosynthesis.txt	87.2	0
human_mitoDB_6_2002_HG-U133A_probes.txt	85.8	0.047
MAP00061_Fatty_acid_biosynthesis_path_1.txt	83.9	0
MAP00253_Tetracycline_biosynthesis.txt	83.9	0
c4_U133_probes.txt	81.5	0.013
c23_U133_probes.txt	80.2	0.002
GO_0005739_HG-U133A_probes.txt	78.0	0.007
MAP00643_Styrene_degradation.txt	77.3	0
MAP00252_Alanine_and_aspartate_metabolism.txt	77.2	0.001
MAP00530_Aminosugars_metabolism.txt	76.9	0
MAP00140_C21_Steroid_hormone_metabolism.txt	73.3	0.001
MAP00290_Valine_leucine_and_isoleucine_biosynthesis.txt	72.7	0.001
MAP00020_Citrate_cycle_TCA_cycle.txt	72.5	0.003
MAP00580_Phospholipid_degradation.txt	70.5	0.001
hypoxia.txt	69.5	0.001
MAP00600_Sphingoglycolipid_metabolism.txt	69.3	0.002
MAP00380_Tryptophan_metabolism.txt	66.6	0.002
MAP00380_Tryptophan_metabolism~.txt	66.6	0
MAP00511_N_Glycan_degradation.txt	64.4	0.001
inflammatory.txt	63.2	0.003
MAP00062_Fatty_acid_biosynthesis_path_2.txt	62.4	0.006
INS_HG-U133A_probes.txt	61.4	0.009
cluster4_LPS_mouse_urinary.txt	60.8	0.002
MAP00910_Nitrogen_metabolism.txt	56.5	0
MAP00650_Butanoate_metabolism.txt	56.3	0.004
cluster2_LPS_mouse_urinary.txt	55.8	0
cluster12_LPS_mouse_urinary.txt	53.3	0.002
MAP00531_Glycosaminoglycan_degradation.txt	52.0	0.008
MAP00220_Urea_cycle_and_metabolism_of_amino_groups.txt	51.3	0
c24_U133_probes.txt	50.5	0.015
MAP00350_Tyrosine_metabolism.txt	50.1	0.001
MAP00500_Starch_and_sucrose_metabolism.txt	50.1	0
MAP00533_Keratan_sulfate_biosynthesis.txt	46.3	0
MAP00520_Nucleotide_sugars_metabolism.txt	44.1	0
MAP00450_Selenoamino_acid_metabolism.txt	44.0	0
cluster3_LPS_mouse_urinary.txt	41.1	0
MAP00052_Galactose_metabolism.txt	39.8	0
MAP00340_Histidine_metabolism.txt	39.4	0.005
cluster9_LPS_mouse_urinary.txt	37.5	0
c22_U133_probes.txt	37.4	0
Hum_Fb_Serum_Chol_Biosyn.txt	35.3	0
MAP00680_Methane_metabolism.txt	34.2	0
c2_U133_probes.txt	33.9	0.092
MAP00512_O_Glycans_biosynthesis.txt	30.1	0
MAP00760_Nicotinate_and_nicotinamide_metabolism.txt	28.6	0.002
MAP00620_Pyruvate_metabolism.txt	26.6	0.001
MAP00770_Pantothenate_and_CoA_biosynthesis.txt	26.2	0
MAP00601_Blood_group_glycolipid_biosynthesis_lact_series.txt	23.2	0.001

MAP00750_Vitamin_B6_metabolism.txt	22.6	0
cluster15_LPS_mouse_urinary.txt	21.2	0.013
MAP00330_Arginine_and_proline_metabolism.txt	20.3	0.002
MAP00562_Inositol_phosphate_metabolism.txt	16.5	0.005
MAP00071_Fatty_acid_metabolism.txt	14.4	0.013
MAP00602_Blood_group_glycolipid_biosynthesis_neolact_series.txt	13.5	0.004
MAP00120_Bile_acid_biosynthesis.txt	12.2	0.013
KET_HG-U133A_probes.txt	9.9	0
MAP00670_One_carbon_pool_by_folate.txt	7.9	0.003
MAP00410_beta_Alanine_metabolism.txt	7.3	0
c25_U133_probes.txt	7.3	0.005
MAP00960_Alkaloid_biosynthesis_II.txt	7.2	0
GLYCOGEN_HG-133A_probes.txt	7.2	0
MAP00271_Methionine_metabolism.txt	7.1	0
MAP00521_Streptomycin_biosynthesis.txt	6.7	0.001
cluster13_LPS_mouse_urinary.txt	4.4	0
MAP00310_Lysine_degradation.txt	4.2	0.001
MAP00830_Retinol_metabolism.txt	4.0	0
MAP00510_N_Glycans_biosynthesis.txt	4.0	0
MAP00640_Propanoate_metabolism.txt	3.3	0.004
cluster7_LPS_mouse_urinary.txt	3.3	0.002
MAP00053_Ascorbate_and_aldarate_metabolism.txt	3.0	0
MAP00631_1_2_Dichloroethane_degradation.txt	3.0	0
MAP00632_Benzoate_degradation.txt	3.0	0.008
MAP00280_Valine_leucine_and_isoleucine_degradation.txt	3.0	0.008
FA_HG-U133A_probes.txt	2.2	0.015
perception_of_pain_sensory_transductuon_of_mechanical_stimulus.txt	1.4	0.001
sensory_perception_of_mechanical_stimulus.txt	1.4	0
MAP00072_Synthesis_and_degradation_of_ketone_bodies.txt	1.4	0.002

Appendix

Down regulated GSEA

Gene Set	ES	P value
MAP00280_Valine_leucine_and_isoleucine_degradation.txt	346.6	0.008
MAP00340_Histidine_metabolism.txt	288.9	0.005
MAP00310_Lysine_degradation.txt	284.6	0
MAP00380_Tryptophan_metabolism.txt	274.2	0.002
MAP00380_Tryptophan_metabolism~.txt	274.2	0
MAP00632_Benzoate_degradation.txt	270.5	0.01
hypoxia.txt	263.3	0
MAP00562_Inositol_phosphate_metabolism.txt	261.8	0.004
MAP00052_Galactose_metabolism.txt	242.5	0
MAP00521_Streptomycin_biosynthesis.txt	239.2	0.001
cluster7_LPS_mouse_urinary.txt	234.1	0
cluster15_LPS_mouse_urinary.txt	232.8	0.014
GLYCOGEN_HG-133A_probes.txt	227.8	0
FA_HG-U133A_probes.txt	225.9	0.011
MAP00072_Synthesis_and_degradation_of_ketone_bodies.txt	219.9	0.001
MAP00410_beta_Alanine_metabolism.txt	216.2	0.003
MAP00071_Fatty_acid_metabolism.txt	210.4	0.027
MAP00510_N_Glycans_biosynthesis.txt	209.1	0
cluster13_LPS_mouse_urinary.txt	207.1	0.001
c23_U133_probes.txt	205.6	0.003
MAP00602_Blood_group_glycolipid_biosynthesis_neolact_series.txt	203.3	0.002
MAP00670_One_carbon_pool_by_folate.txt	202.3	0.001
Hum_Fb_Serum_Chol_Biosyn.txt	200.3	0
MAP00640_Propanoate_metabolism.txt	198.9	0.005
c22_U133_probes.txt	196.6	0
MAP00140_C21_Steroid_hormone_metabolism.txt	193.0	0.001
MAP00650_Butanoate_metabolism.txt	183.0	0.006
MAP00120_Bile_acid_biosynthesis.txt	179.0	0.008
MAP00910_Nitrogen_metabolism.txt	176.5	0.001
MAP00750_Vitamin_B6_metabolism.txt	174.1	0.001
MAP00830_Retinol_metabolism.txt	172.3	0
c25_U133_probes.txt	171.3	0.003
MAP00053_Ascorbate_and_aldarate_metabolism.txt	166.2	0
MAP00631_1_2_Dichloroethane_degradation.txt	166.2	0
inflammatory.txt	162.9	0.004
MAP00220_Urea_cycle_and_metabolism_of_amino_groups.txt	162.1	0
MAP00760_Nicotinate_and_nicotinamide_metabolism.txt	160.5	0.001
MAP00360_Phenylalanine_metabolism.txt	159.4	0
c34_U133_probes.txt	158.8	0.003
c2_U133_probes.txt	154.3	0.075
cluster9_LPS_mouse_urinary.txt	153.7	0
MAP00580_Phospholipid_degradation.txt	153.6	0.003
GO_0005739_HG-U133A_probes.txt	152.7	0.008
MAP00601_Blood_group_glycolipid_biosynthesis_lact_series.txt	151.7	0.002
MAP00603_Globoside_metabolism.txt	151.7	0.002

cluster12_LPS_mouse_urinary.txt	151.5	0.003
MAP00770_Pantothenate_and_CoA_biosynthesis.txt	150.2	0
MAP00193_ATP_synthesis.txt	149.7	0
MAP03070_Type_III_secretion_system.txt	149.7	0
c24_U133_probes.txt	148.6	0.01
MAP00271_Methionine_metabolism.txt	146.7	0
MAP00330_Arginine_and_proline_metabolism.txt	144.5	0.003
MAP00533_Keratan_sulfate_biosynthesis.txt	143.7	0
MAP00511_N_Glycan_degradation.txt	142.3	0
KET_HG-U133A_probes.txt	141.0	0.001
MAP00061_Fatty_acid_biosynthesis_path_1.txt	138.4	0
MAP00253_Tetracycline_biosynthesis.txt	138.4	0
MAP00960_Alkaloid_biosynthesis_II.txt	138.3	0
MAP00062_Fatty_acid_biosynthesis_path_2.txt	137.6	0.006
human_mitoDB_6_2002_HG-U133A_probes.txt	137.2	0.039
cluster2_LPS_mouse_urinary.txt	134.0	0
Hum_Fb_Serum_Coag_Hemo.txt	132.6	0.003
MAP00600_Sphingoglycolipid_metabolism.txt	130.2	0
INS_HG-U133A_probes.txt	126.9	0.009
MAP00520_Nucleotide_sugars_metabolism.txt	125.4	0
MAP00531_Glycosaminoglycan_degradation.txt	123.5	0.002
cluster4_LPS_mouse_urinary.txt	123.2	0
MAP03030_DNA_polymerase.txt	121.8	0.001
cluster3_LPS_mouse_urinary.txt	121.7	0.001
c1_U133_probes.txt	120.4	0.047
MAP00530_Aminosugars_metabolism.txt	120.3	0
c4_U133_probes.txt	118.4	0.014
perception_of_pain_sensory_transductuon_of_mechanical_stimulus.txt	116.9	0
sensory_perception_of_mechanical_stimulus.txt	116.9	0
MAP00195_Photosynthesis.txt	116.3	0.001
MAP00350_Tyrosine_metabolism.txt	115.5	0.001
MAP00920_Sulfur_metabolism.txt	113.0	0.004
MAP00020_Citrate_cycle_TCA_cycle.txt	110.6	0.002
MAP00450_Selenoamino_acid_metabolism.txt	109.8	0
MAP00532_Chondroitin_Heparan_sulfate_biosynthesis.txt	106.4	0
MAP00512_O_Glycans_biosynthesis.txt	105.7	0.001
MAP00290_Valine_leucine_and_isoleucine_biosynthesis.txt	105.5	0
MAP00790_Folate_biosynthesis.txt	105.1	0
MAP00500_Starch_and_sucrose_metabolism.txt	102.3	0
MAP00100_Sterol_biosynthesis.txt	100.9	0.03
MAP00522_Erythromycin_biosynthesis.txt	99.8	0
MAP00620_Pyruvate_metabolism.txt	98.4	0
MAP00272_Cysteine_metabolism.txt	98.1	0
c15_U133_probes.txt	97.0	0
MAP00010_Glycolysis_Gluconeogenesis.txt	96.1	0
MAP00150_Androgen_and_estrogen_metabolism.txt	95.9	0.001
MAP00040_Pentose_and_glucuronate_interconversions.txt	95.5	0
MAP00720_Reductive_carboxylate_cycle_CO2_fixation.txt	94.9	0
MAP00680_Methane_metabolism.txt	93.1	0
MAP00252_Alanine_and_aspartate_metabolism.txt	92.4	0.001

cluster10_LPS_mouse_urinary.txt	92.0	0
MAP00625_Tetrachloroethene_degradation.txt	90.1	0
MAP03020_RNA_polymerase.txt	89.2	0
c7_U133_probes.txt	88.6	0.096
c27_U133_probes.txt	87.4	0
MAP00630_Glyoxylate_and_dicarboxylate_metabolism.txt	83.9	0
c9_U133_probes.txt	82.7	0.004
MAP00900_Terpenoid_biosynthesis.txt	82.5	0.001
MAP00471_D_Glutamine_and_D_glutamate_metabolism.txt	81.0	0
MAP00860_Porphyrin_and_chlorophyll_metabolism.txt	80.8	0
TCA_HG-U133A_probes.txt	78.1	0.006
cluster14_LPS_mouse_urinary.txt	76.5	0
c29_U133_probes.txt	76.4	0
MAP00051_Fructose_and_mannose_metabolism.txt	76.0	0
mitochondr_HG-U133A_probes.txt	75.0	0.016
c35_U133_probes.txt	74.4	0
MAP00643_Styrene_degradation.txt	74.4	0
MAP00260_Glycine_serine_and_threonine_metabolism.txt	74.4	0.001
inflammatory_mechanical.txt	74.3	0
c21_U133_probes.txt	72.9	0.003
MAP00780_Biotin_metabolism.txt	71.9	0
MAP00400_Phenylalanine_tyrosine_and_tryptophan_biosynthesis.txt	71.3	0
c28_U133_probes.txt	69.0	0.001
MAP00430_Taurine_and_hypotaurine_metabolism.txt	68.5	0.004
cluster1_LPS_mouse_urinary.txt	66.6	0
MAP00300_Lysine_biosynthesis.txt	62.2	0.005
c19_U133_probes.txt	59.5	0
MAP00240_Pyrimidine_metabolism.txt	59.5	0.005
MAP00130_Ubiquinone_biosynthesis.txt	58.7	0
c10_U133_probes.txt	56.1	0.01
MAP00230_Purine_metabolism.txt	55.2	0.004
MAP00472_D_Arginine_and_D_ornithine_metabolism.txt	54.1	0
c32_U133_probes.txt	54.1	0
MAP00740_Riboflavin_metabolism.txt	52.6	0
c30_U133_probes.txt	51.9	0.001
MAP00590_Prostaglandin_and_leukotriene_metabolism.txt	51.9	0
c0_U133_probes.txt	51.3	0.042
cluster6_LPS_mouse_urinary.txt	50.2	0
ROS_HG-U133A_probes.txt	49.6	0
MAP00190_Oxidative_phosphorylation.txt	49.5	0
c8_U133_probes.txt	47.7	0.027
MAP00561_Glycerolipid_metabolism.txt	47.5	0.001
cluster11_LPS_mouse_urinary.txt	47.2	0.009
c3_U133_probes.txt	46.7	0
MAP00480_Glutathione_metabolism.txt	46.4	0
MAP00950_Alkaloid_biosynthesis_I.txt	46.0	0
c20_U133_probes.txt	44.9	0.002
c12_U133_probes.txt	44.1	0
cluster5_LPS_mouse_urinary.txt	43.5	0.001
c17_U133_probes.txt	41.8	0.001

c6_U133_probes.txt	41.1	0.025
c31_U133_probes.txt	38.1	0
MAP00570_Sphingophospholipid_biosynthesis.txt	37.8	0
c16_U133_probes.txt	32.0	0.004
GLUCO_HG-U133A_probes.txt	26.5	0
MAP00970_Aminoacyl_tRNA_biosynthesis.txt	26.3	0.013
c33_U133_probes.txt	25.4	0
MAP00940_Flavonoids_stilbene_and_lignin_biosynthesis.txt	21.8	0
c13_U133_probes.txt	20.6	0
PYR_HG-U133A_probes.txt	19.6	0.003
c5_U133_probes.txt	19.6	0.002
c14_U133_probes.txt	18.3	0.002
MAP00361_gamma_Hexachlorocyclohexane_degradation.txt	16.1	0.004
OXPHOS_HG-U133A_probes.txt	15.2	0.131
MAP00710_Carbon_fixation.txt	13.9	0.002
c18_U133_probes.txt	13.8	0.035
GLYCOL_HG-U133A_probes.txt	13.5	0
MAP00030_Pentose_phosphate_pathway.txt	11.3	0.001
c11_U133_probes.txt	9.5	0.001
Hum_Fb_Serum_EarlyTF.txt	7.2	0.003
c26_U133_probes.txt	7.0	0.029
MAP00251_Glutamate_metabolism.txt	5.8	0.001
ECM_constituent_conferring_elasticity.txt	5.1	0
ECM_structural_constituent_conferring_compression_resistance.txt	5.1	0
MAP00440_Aminophosphonate_metabolism.txt	4.8	0
MAP00460_Cyanoamino_acid_metabolism.txt	4.8	0
response_to_mechanical_stimulus.txt	4.4	0.108
MAP00031_Inositol_metabolism.txt	4.3	0

Appendix

Microchip preparation and RNA extraction protocol (from Travis Burleson at the Children's Hospital microarray facility)

Introduction

Microarray technology has greatly enhanced and accelerated the ability to analyze relevant changed that occur in biological samples. Affymetrix has long been the gold standard in this field. They are currently producing their most recent chip to analyze the entire human genome on a single array. The GeneChip Human Genome U133 Plus 2.0 array provides comprehensive coverage of the transcribed human genome. It consists of over 54,000 probe sets and 1,300,000 distinct oligonucleotide features. The sequences for the probe sets were selected from GenBank, dbEST, and RefSeq databases, while the sequence clusters were created from the UniGene database.

The scientists in the Microarray Facility at Children's Hospital Boston are experts in Affymetrix technology. With the most recent scanners and analysis programs, as well as years of experience, they provide excellent an excellent group to have biological samples analyzed.

Equipment and Capacity

The Microarray Core Facility owns two GeneChip Hybridization Ovens, three GeneChip Fluidics Station 450s, and a GeneChip Scanner 3000 with Autoloader. At capacity, the core can extract samples in batches of 6 and perform the reverse transcription/invitro transcription reactions in batches of 6 to 12. Up to 24 chips can be analyzed per week.

Tissue Criteria

The single most important determinant in the success of a GeneChip assay is the quality of the RNA. Differential degradation of RNA can lead to erroneous conclusions about both the relative and absolute mRNA levels in the specimens. The quality of RNA is

directly related to proper tissue handling procedures which will be out of the control of the Microarray Core Facility. We expect the tissue samples will meet the following:

1. The tissue samples will have been frozen in liquid nitrogen within one hour of being taken.
2. The tissue samples will be predominantly skeletal muscle.
3. The size of each tissue sample will be 70mg or larger.
4. All tissue samples will be shipped on dry ice and not have undergone multiple freeze/thaw cycles.

Upon arrival, we will store the tissue samples in a -80C freezer until being extracted. If a tissue passes visual inspection, it will undergo total RNA isolation (outlined in the RNA isolation section). After the total RNA is isolated, it will be analyzed on a 1% agarose gel for quality purposes. The gel will be employed to determine if the 28S:18S ribosomal bands have a ratio of 2:1. If the gel fails to meet our criteria by not having a 2:1 ratio, or by containing any other anomalies, the sample will be re-extracted. Following this second extraction, the total RNA sample will be analyzed again on a 1% agarose gel.

RNA Isolation

Patient tissue samples weighing 70-120 mg will be homogenized in 1 ml of ice cold Trizol (Life Technologies, Rockville, MD) for 30 seconds at room temperature then centrifuged at 10° C to remove cell debris. DNA and proteins will be removed by chloroform:isoamyl alcohol (49:1) extraction, and Isopropanol, sodium citrate (0.8M) and sodium chloride (1.2M) will be used for RNA precipitation. RNA concentration is measured using a spectrophotometer and RNA quality will be evaluated by running 1 μg of RNA on a 2% agarose gel.

Target (labeled cRNA) Preparation

Good quality total RNA is the necessary starting material to obtain labeled RNA. Approximately 7 μg of total RNA will be used for cDNA synthesis (Superscript II kit, Life Technology). Biotin labeled cRNA is produced through in vitro transcription using

the cDNA of the previous reaction (BioArray High Yield RNA transcript Labeling kit, Enzo, Farmingdale, NY).

The biotin-labeled RNA will be analyzed on a spectrophotometer to determine concentration and purity. Final RNA output should be greater that 20ug and have a 260/280 ratio of 1.8-2.2. 20ug of the cRNA is then fragmented in the presence of heat and Mg^{++}, before a hybridization cocktail is created.

Hybridization Cocktail

Hybridization Coctails will be prepared according to the procedure outlined by Affymetrix. In brief, 15ul of fragmented biotin-labled cRNA, Control Oligonucleotide B2, Hybridization Controls (bioB, bioC, bioD, and cre), BSA, Herring Sperm DNA, and buffer will be mixed.

Array Hybridization, Wash, Stain and Scan (HWSS):

200ul of the Hybridization Coctail is hybridized to an array for 16 hours at 45^0 C and 60RPM. The array is then washed and stained in a GeneChip Fluidics Station 450. The staining protocol is a antibody amplification. The first stain with streptavidin-phycoerythrin is followed by an antibody amplification with Goat IgG and biotinylated antibody. The final stain is with streptavidin-phycoerythin.

The washed and stained chip will be scanned on an Affymetrix GeneChip Scanner 3000.

Array Data Analysis and Quality Control Standards

Scanned chip images will be analyzed using GCOS v1.1 and analyzed for quality. The parameters that determine good quality data include: the 3'/5' ratio of housekeeping genes B-Actin and GAPDH; Background values; Noise (Q) values; Scaling Factor values; and Spiked Control values. We have outlined values that the data should adhere to. The sample data should meet outlined values, but deviation in a quality control value may not cause the sample to be discarded; rather the global quality will also be taken into

account. The Core will contact Wyeth about any data in question and make suggestions about whether the sample is good or bad.

3'/5' ratio of housekeeping genes: This is a measure of the efficiency of the cDNA synthesis reaction. Reverse transcriptase synthesizes cDNA starting from the 3'-end of an mRNA and ending at the 5'-end. All Affymetrix arrays contain probes for the regions corresponding to 3', middle and 5'-end of the house keeping genes such as GAPDH and B-Actin. The ratio of signal intensity for 3' probes to that from 5' probes provides a measure of the number of cDNA synthesis reactions that went to completion (full length cDNA is synthesized). An ideal ratio would be 1 whereas a higher value indicates that many cDNAs were started but did not go to completion. The 3'/5' ratio for these housekeeping genes should be at most 3. If the ratio is above 3, the sample will be resubmitted to Wyeth to ascertain if the sample should be used.

Using a reference experiment of 63 muscle biopsy samples extracted and labeled according to the above outlined protocol and analyzed on U133A chips, we observe a B-Actin average value of 1.45, and a GAPDH average value of 1.5. Although these values are for U133A chips, we expect similar values using the U133 Plus 2.0 arrays.

Background values with standard deviation. Background value is a measure of the signal intensity caused by autofluorescence of the array surface as well as nonspecific binding of target or stain molecules (SAPE). The background values for all the arrays in one experiment should be very similar to each other; otherwise comparison data may not be accurate. Non-specific binding causes a low signal to noise ratio, which means that genes for transcripts present at very low levels in the sample may incorrectly be called absent. Thus, high background creates an overall loss of sensitivity in the experiment. Any samples with a background value over 120 will be resubmitted to Wyeth to ascertain if the data should be used.

Using a reference experiment of 63 muscle biopsy samples extracted and labeled according to the above outlined protocol and analyzed on U133A chips, we observe a

Background average value of 79.9. Although these values are for U133A chips, we expect similar values using the U133 Plus 2.0 arrays.

Noise (Q) value. Noise (Q value) results from small variations in the digitized signal observed by the scanner as it samples the probe array's surface. It is measured by examining pixel-to-pixel variations in background intensities. The noise value for all the arrays in one experiment should be very similar to each other. Any samples with a Noise (Q) value over 5 will be resubmitted to Wyeth to ascertain if the data should be used.

Using a reference experiment of 63 muscle biopsy samples extracted and labeled according to the above outlined protocol and analyzed on U133A chips, we observe a Noise (Q) average value of 3.02. Although these values are for U133A chips, we expect similar values using the U133 Plus 2.0 arrays.

Scaling factor. The scaling factor provides a measure of the brightness of the array. The "brightness" (image intensity) varies from array to array. Non-biological factors (amount and quality of the cRNA, amount of stain or other experimental variation) can contribute to the overall variability in hybridization intensities. In order to reliably compare data from multiple arrays, it is essential that the intensity of the arrays be brought to the same level. Scaling is a mathematical technique used by the Gene Chip Operating Software (GCOS) to minimize differences in overall signal intensities between two or more arrays thus allowing for more reliable detection of biologically relevant changes in the same sample. GCOS calculates the overall intensity of an array by averaging the intensity values of every probe set on the array with the exception of the top and bottom 2% of the probe set intensities. The average intensity of the array is then multiplied by the Scaling factor to bring it to an arbitrary Target Intensity value (usually 1500) set by the user. Thus, scaling allows a number of experiments to become normalized to one Target Intensity, allowing comparison between any two experiments. In a particular set of experiment, the Scaling Factor value for all the arrays should be very close to each other (within three-fold of each other). Any samples with a scaling value over 12 will be resubmitted to Wyeth to ascertain if the data should be used.

Using a reference experiment of 63 muscle biopsy samples extracted and labeled according to the above outlined protocol and analyzed on U133A chips, we observe a scaling average value of 9.44. Although these values are for U133A chips, we expect similar values using the U133 Plus 2.0 arrays.

Presence of spiked control cRNAs: Bio-B, C, D and CRE serve as a controls for hybridization and are spiked at the following concentrations: BioB: 1.5 pM, BioC: 5.0 pM, BioD: 25.0 pM, BioCRE: 100 pM. We specifically look at the average difference values which should be present in increasing amounts, B being the least and CRE the highest.

Microarray expression array protocol:

Whole frozen tissue samples are mixed with TRIzol reagent and homogenized completely. This mixture is centrifuged for ten minutes at 4 c and then incubated at room temperature for five minutes to dissociate nucleoprotein complexes. 200 μL of chloroform is added, the sample is shaken to homogenize the solution, and is incubated at room temperature for two minutes. The sample is then spun again in a centrifuge for 15 minutes at 4 c. The aqueous layer is extracted and to it is added 250 μL of isopropyl alcohol and 250 μL of a salt solution containing sodium citrate and sodium chloride. The sample is mixed thoroughly, incubated at room temperature for ten minutes and spun in a centrifuge for ten minutes at 4 c. The supernatant is removed, the pellet is washed with 75% EtOH, and then resuspended in RNAse-free water.

Following extraction, the RNA is quantified and checked for purity on a spectrophotometer. If the RNA is of sufficient quality and quantity, it then is used for the synthesis of cDNA. 7 μg of the extracted RNA is used in this reaction. We use the *SuperScript Double Stranded cDNA Synthesis* kit from Invitrogen for this reaction. This kit contains all enzymes and reagents needed for the synthesis of double stranded cDNA. This reaction takes between 3 and 4 hours (depending on set-up and clean-up time). The product of this reaction is cleaned in a standard phenol-chloroform type extraction using a phase-lock gel. To the resulting solution containing the double stranded cDNA is added

2.5 volumes of 100% EtOH and 0.5 volumes of 7.5M ammonium acetate. This solution is incubated at –20 c overnight to allow the cDNA to precipitate.

The precipitation is spun in a centrifuge and the pellet is cleaned in EtOH and resuspended in Rnase-free water. This product is used in the next step: Invitro Translation. For this reaction, we use the *Bioarray* biotin labeling kit from Enzo. This kit synthesizes single stranded cRNA from the cDNA product produced in the last step. This kit also incorporates the biotin labeled nucleotides necessary for hybridization. This kit contains all necessary enzymes and reagents for the reaction. This reaction takes approximately 4.5 hours (including preparation), which is followed by a clean-up using the *RNeasy* kit supplied by Qiagen. This kit contains all necessary buffers and columns for the process. This step takes approximately 0.5 hours. The resulting solution containing the labeled cRNA is subjected to the same precipitation process as detailed above.

Again, the precipitation is spun, washed, and resuspended. The cRNA is quantified and checked for purity on a spectrophotometer. Approximately 20 μg of the cRNA is used in the fragmentation step. The cRNA is added to a solution containing magnesium acetate and potassium acetate. By a process of metal induced hydrolysis, the cRNA is broken down into 35 – 200 bp fragments. This process takes 35 minutes and is performed at 95 c. Following the fragmentation, other components, including BSA, Herring Sperm DNA, and spike and housekeeping controls, are added to the cRNA to form a 'Hybridization Cocktail'. This cocktail is then injected onto the array, hybridized for 16 hours, and scanned. The hybridization cocktails can be used up to 5 times and can last for up to a year if stored at –80 c.

Appendix

Validation of the FEA using order of magnitude analysis

Order of Magnitude Analysis

The atmosphere imposes a uniform pressure of about 1 bar (or 10^5 Pa). The force applied to a system is equal to the differential pressure (gage pressure) multiplied by the surface area over which the pressure acts.

Think of a vacuum cup applied to a surface. Let us start evacuating the air from the cup. If we have a good seal, no air will be allowed to come into the cup to replace the air lost. Thus, because we have reduced the number of molecules in the system—and assuming the temperature is allowed to stay constant, and that the cup is rigid such that its volume is fixed, we see from the perfect gas law that the pressure in the system will fall:

$P_1V=n_1RT$ ...(1)

The '1' subscript refers to conditions before the evacuation of molecules. No subscript on a variable implies that that variable is held fixed. Next after the decrease in pressure,

$P_2V=n_2RT$ ...(2)

The '2' subscript on variables refers to conditions after the evacuation of pressure. Dividing (1) by (2), we obtain $\frac{P_1}{P_2}=\frac{n_1}{n_2}$, or $P_2=\frac{n_2}{n_1}P_1$, which says that as n_2 goes down the pressure P_2 decreases.

Force is nothing but the pressure times the area. Now the pressure always acts normal to a surface. Since the pressure on the inside of the cup is lower than atmospheric pressure, a lower force acts on this side. If the cup is rigid, as it is in this example, then the difference in forces will be countered by the stresses set up in the cup wall. If however, the system is allowed to be flexible, then the difference in pressures (and hence forces) will cause the system to collapse, until we come to an equilibrium.

In the equilibrium state, the difference in pressure still exists. This difference in pressure is taken up by the stresses set up in the wound itself (as well as stresses in the material making up the vac sponge). We thus see that the pressure differential causes the vac sponge to collapse. However, the reduction in volume is not enough to cause the pressure to become atmospheric inside the seal. Furthermore, the material of the vac sponge itself is much less compliant than the wound bed (once the sponge has compressed—before the compression, the overall vac sponge compliance is higher than the wound). Thus, the wound now locally (between the pores that define the vac sponge) rises.

Now let's see what kinds of stresses are set up—assuming a thin walled system for the wound. This is a rough approximation, because the wound is not a thin walled vessel. The thin wall approximation allows us to assume that the stresses in the wall are uniform, so the results we obtain from this theoretical framework give us an order of magnitude understanding and validation of the finite element results.

For a thin membrane acted on by an internal pressure, we equate the pressure forces acting over the inside surface to the stresses acting over the wall thickness. Thus, $\sigma 2\pi rt = \pi r^2 P_i$, where P_i is the pressure differential applied to the wound, and r is the radius of the pore of the vac sponge. σ is the stress in the wound bed. Solving for σ, we obtain $\sigma = \frac{P_i r}{2t}$, and putting in the values for our pressure—and assuming a thickness of about 1mm, we obtain for the stress an order of magnitude of $\sigma = 9{,}000Pa$. Now we use the modulus of elasticity of 50,000Pa to obtain the strain. This is, $\varepsilon = \frac{\sigma}{E} = 18\%$, and this is on the order of magnitude of the strain we obtained in our analyses.

References

1. Lott-Crumpler, D.A. and H.R. Chaudhry, *Optimal patterns for suturing wounds of complex shapes to foster healing.* J Biomech, 2001. **34**(1): p. 51-8.
2. Saxena, V., et al., *Vacuum-assisted closure: microdeformations of wounds and cell proliferation.* Plast Reconstr Surg, 2004. **114**(5): p. 1086-96; discussion 1097-8.
3. Ilizarov, G.A., *The tension-stress effect on the genesis and growth of tissues: Part II. The influence of the rate and frequency of distraction.* Clin Orthop Relat Res, 1989(239): p. 263-85.
4. Ilizarov, G.A., *The tension-stress effect on the genesis and growth of tissues. Part I. The influence of stability of fixation and soft-tissue preservation.* Clin Orthop Relat Res, 1989(238): p. 249-81.
5. Huang, S. and D.E. Ingber, *The structural and mechanical complexity of cell-growth control.* Nat Cell Biol, 1999. **1**(5): p. E131-8.
6. Huang, S., C.S. Chen, and D.E. Ingber, *Control of cyclin D1, p27(Kip1), and cell cycle progression in human capillary endothelial cells by cell shape and cytoskeletal tension.* Mol Biol Cell, 1998. **9**(11): p. 3179-93.
7. Chen, C.S., et al., *Geometric control of cell life and death.* Science, 1997. **276**(5317): p. 1425-8.
8. Mochitate, K., P. Pawelek, and F. Grinnell, *Stress relaxation of contracted collagen gels: disruption of actin filament bundles, release of cell surface fibronectin, and down-regulation of DNA and protein synthesis.* Exp Cell Res, 1991. **193**(1): p. 198-207.
9. Korff, T. and H.G. Augustin, *Tensional forces in fibrillar extracellular matrices control directional capillary sprouting.* J Cell Sci, 1999. **112 (Pt 19)**: p. 3249-58.
10. Nagel, T., et al., *Vascular endothelial cells respond to spatial gradients in fluid shear stress by enhanced activation of transcription factors.* Arterioscler Thromb Vasc Biol, 1999. **19**(8): p. 1825-34.
11. Gimbrone, M.A., Jr., T. Nagel, and J.N. Topper, *Biomechanical activation: an emerging paradigm in endothelial adhesion biology.* J Clin Invest, 1997. **100**(11 Suppl): p. S61-5.
12. Tardy, Y., et al., *Shear stress gradients remodel endothelial monolayers in vitro via a cell proliferation-migration-loss cycle.* Arterioscler Thromb Vasc Biol, 1997. **17**(11): p. 3102-6.
13. Miyajima, K., *[Effects of periodic tension on osteoblast-like cells for cell differentiation and alkaline phosphatase activity].* Nippon Kyosei Shika Gakkai Zasshi, 1990. **49**(3): p. 226-36.
14. Akhyari, P., et al., *Mechanical stretch regimen enhances the formation of bioengineered autologous cardiac muscle grafts.* Circulation, 2002. **106**(12 Suppl 1): p. I137-42.
15. Joseph, E., et al., *A prospective randomized trial of vacuum-assisted closure versus standard therapy of chronic nonhealing wounds.* Wounds-a Compendium of Clinical Research and Practice, 2000. **12**(3): p. 60-67.

16. McCallon, S.K., et al., *Vacuum-assisted closure versus saline-moistened gauze in the healing of postoperative diabetic foot wounds.* Ostomy Wound Manage, 2000. **46**(8): p. 28-32, 34.
17. Argenta, L.C. and M.J. Morykwas, *Vacuum-assisted closure: A new method for wound control and treatment: Clinical experience.* Annals of Plastic Surgery, 1997. **38**(6): p. 563-576.
18. Webb, L.X., *New techniques in wound management: vacuum-assisted wound closure.* J Am Acad Orthop Surg, 2002. **10**(5): p. 303-11.
19. Morykwas, M.J., et al., *Vacuum-assisted closure: a new method for wound control and treatment: animal studies and basic foundation.* Ann Plast Surg, 1997. **38**(6): p. 553-62.
20. Argenta, L.C. and M.J. Morykwas, *Vacuum-assisted closure: a new method for wound control and treatment: clinical experience.* Ann Plast Surg, 1997. **38**(6): p. 563-76; discussion 577.
21. Armstrong, D.G., et al., *Outcomes of subatmospheric pressure dressing therapy on wounds of the diabetic foot.* Ostomy Wound Manage, 2002. **48**(4): p. 64-8.
22. Clare, M.P., et al., *Experience with the vacuum assisted closure negative pressure technique in the treatment of non-healing diabetic and dysvascular wounds.* Foot Ankle Int, 2002. **23**(10): p. 896-901.
23. Kinetic Concepts Incorporated, K., San Antonio, *http://www.kci1.com/.*
24. devicelink, m., http://www.devicelink.com/mx/archive/05/03/coverbx.html.
25. Zoysa, D.d., Bret Jones, and D. Clayton, *Medical Supplies and Devices, Industry outlook, Advanced Wound Care: Beyond Band-Aids.* march 2005, SG Cowen and Co.
26. http://sprojects.mmi.mcgill.ca/dermatology/.
27. http://www.homestead.com/doctorderm/skinanatomy.html.
28. Davidson, C.R.H.a.J.M., *Biochemical and cellular aspects of wound repair*, in *Mammalian Development*, P. Lonai, Editor. 1996, Harwood Academic Publishers: Amsterdam. p. 293-320.
29. http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/C/Clotting.html.
30. Lo, C.M., et al., *Cell movement is guided by the rigidity of the substrate.* Biophysical Journal, 2000. **79**(1): p. 144-152.
31. Gottrup, F., M.S. Agren, and T. Karlsmark, *Models for use in wound healing research: A survey focusing on in vitro and in vivo adult soft tissue.* Wound Repair and Regeneration, 2000. **8**(2): p. 83-96.
32. Dipietro, L.A. and A.L. Burns, eds. *Wound healing, methods and protocols.* 2003, Humana Press: Totowa, New Jersey.
33. Lindblad, W.J., *Animal models in wound healing research: Do we need more?* Wound Repair and Regeneration, 2000. **8**(2): p. 81-82.
34. Hollander, D.A., et al., *Standardized qualitative evaluation of scar tissue properties in an animal wound healing model.* Wound Repair and Regeneration, 2003. **11**(2): p. 150-157.
35. Murray, J.D., *Mathematical Biology II: Spatial models and biomedical applications.* 3 ed. 2003: Springer-Verlag.
36. Wang, N., J.P. Butler, and D.E. Ingber, *Mechanotransduction across the Cell-Surface and through the Cytoskeleton.* Science, 1993. **260**(5111): p. 1124-1127.

37. Wakatsuki, T., et al., *Effects of cytochalasin D and latrunculin B on mechanical properties of cells.* J Cell Sci, 2001. **114**(Pt 5): p. 1025-36.
38. Dembo, M. and Y.L. Wang, *Stresses at the cell-to-substrate interface during locomotion of fibroblasts.* Biophysical Journal, 1999. **76**(4): p. 2307-2316.
39. Wang, N., et al., *Micropatterning tractional forces in living cells.* Cell Motility and the Cytoskeleton, 2002. **52**(2): p. 97-106.
40. Waters, C.M., et al., *Cellular biomechanics in the lung.* American Journal of Physiology-Lung Cellular and Molecular Physiology, 2002. **283**(3): p. L503-L509.
41. D'Addario, M., et al., *Regulation of tension-induced mechanotranscriptional signals by the microtubule network in fibroblasts.* J Biol Chem, 2003. **278**(52): p. 53090-7.
42. Chiquet, M., et al., *How do fibroblasts translate mechanical signals into changes in extracellular matrix production?* Matrix Biol, 2003. **22**(1): p. 73-80.
43. Kraiss, L.W., et al., *Fluid flow activates a regulator of translation, p70/p85 S6 kinase, in human endothelial cells.* Am J Physiol Heart Circ Physiol, 2000. **278**(5): p. H1537-44.
44. Chien, S., S. Li, and Y.J. Shyy, *Effects of mechanical forces on signal transduction and gene expression in endothelial cells.* Hypertension, 1998. **31**(1 Pt 2): p. 162-9.
45. Ji, J.Y., H. Jing, and S.L. Diamond, *Shear stress causes nuclear localization of endothelial glucocorticoid receptor and expression from the GRE promoter.* Circ Res, 2003. **92**(3): p. 279-85.
46. Remmler, D., et al., *Presurgical finite element analysis from routine computed tomography studies for craniofacial distraction: II. An engineering prediction model for gradual correction of asymmetric skull deformities.* Plast Reconstr Surg, 1998. **102**(5): p. 1395-404.
47. Orgill, D.P., et al., *A finite-element model predicts thermal damage in cutaneous contact burns.* J Burn Care Rehabil, 1998. **19**(3): p. 203-9.
48. Mizunuma, M., et al., *Can dog-ear formation be decreased when an S-shaped skin resection is used instead of a spindle skin resection? A three-dimensional analysis of skin surgery techniques using the finite element method.* Plast Reconstr Surg, 2000. **106**(4): p. 845-8; discussion 849-51.
49. Fung, Y.C., *Biomechanics : mechanical properties of living tissues.* 2nd ed. 1993, New York: Springer-Verlag. xviii, 568.
50. Bjerring, P., *Skin elasticity measured by dynamic admittance. A new technique for mechanical measurements in patients with scleroderma.* Acta Derm Venereol Suppl (Stockh), 1985. **120**: p. 83-7.
51. Ingber, D.E., *Fibronectin controls capillary endothelial cell growth by modulating cell shape.* Proc Natl Acad Sci U S A, 1990. **87**(9): p. 3579-83.
52. Ingber, D.E., et al., *Cell shape, cytoskeletal mechanics, and cell cycle control in angiogenesis.* J Biomech, 1995. **28**(12): p. 1471-84.
53. Costa, K.D., E.J. Lee, and J.W. Holmes, *Creating alignment and anisotropy in engineered heart tissue: role of boundary conditions in a model three-dimensional culture system.* Tissue Eng, 2003. **9**(4): p. 567-77.

54. Mootha, V.K., et al., *PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes.* Nat Genet, 2003. **34**(3): p. 267-73.
55. Mootha, V.K., *http://www.broad.mit.edu/mpg/oxphos/.*
56. GeneOntology, *http://www.geneontology.org.*
57. Kohane, I.S., A.T. Kho, and A.J. Butte, *Microarrays for an integrative genomics.* Computational molecular biology. 2003, Cambridge, Mass.: MIT Press. xviii, 306.
58. Good, P.I., *Permutation tests : a practical guide to resampling methods for testing hypotheses.* 2nd ed. Springer series in statistics. 2000, New York: Springer. xvi, 270.
59. Saban, M.R., et al., *Time course of LPS-induced gene expression in a mouse model of genitourinary inflammation.* Physiol Genomics, 2001. **5**(3): p. 147-60.
60. Kohonen, T., *Self-organizing maps.* 3rd ed. Springer series in information sciences ; 30. 2001, Berlin ; New York: Springer. xx, 501.
61. Broad, *http://www.broad.mit.edu.*
62. BiologyOnline, *http://www.biology-online.org/dictionary/hypoxia.*
63. Lu, S.Y., et al., *Inhibition of hypoxia-induced proliferation and collagen synthesis by vasonatrin peptide in cultured rat pulmonary artery smooth muscle cells.* Life Sci, 2005. **77**(1): p. 28-38.
64. An, S.S., et al., *Hypoxia alters biophysical properties of endothelial cells via p38 MAPK- and Rho kinase-dependent pathways.* Am J Physiol Cell Physiol, 2005.
65. Szpaderska, A.M., et al., *Distinct patterns of angiogenesis in oral and skin wounds.* J Dent Res, 2005. **84**(4): p. 309-14.
66. Sahota, P.S., et al., *Approaches to improve angiogenesis in tissue-engineered skin.* Wound Repair Regen, 2004. **12**(6): p. 635-42.
67. Kondo, Y., et al., *Over expression of hypoxia-inducible factor-1alpha in renal and bladder cancer cells increases tumorigenic potency.* J Urol, 2005. **173**(5): p. 1762-6.
68. Schlueter, C., et al., *Angiogenetic signaling through hypoxia: HMGB1: an angiogenetic switch molecule.* Am J Pathol, 2005. **166**(4): p. 1259-63.
69. Powell, J., *Vascular Biology.* http://intl.elsevierhealth.com/e-books/pdf/787.pdf. 2003: Elsevier.
70. Tagawa, T., et al., *Role of nitric oxide in reactive hyperemia in human forearm vessels.* Circulation, 1994. **90**(5): p. 2285-90.
71. Zhao, J.L., et al., *Bioactive nitric oxide concentration does not increase during reactive hyperemia in human skin.* J Appl Physiol, 2004. **96**(2): p. 628-32.
72. Romero-Graillet, C., et al., *Nitric oxide produced by ultraviolet-irradiated keratinocytes stimulates melanogenesis.* J Clin Invest, 1997. **99**(4): p. 635-42.
73. Sikorski, E.M., et al., *The story so far: Molecular regulation of the heme oxygenase-1 gene in renal injury.* Am J Physiol Renal Physiol, 2004. **286**(3): p. F425-41.
74. Turcanu, V., M. Dhouib, and P. Poindron, *Heme oxygenase inhibits nitric oxide synthase by degrading heme: a negative feedback regulation mechanism for nitric oxide production.* Transplant Proc, 1998. **30**(8): p. 4184-5.

75. Ishizaka, N., et al., *Heme oxygenase-1 is upregulated in the rat heart in response to chronic administration of angiotensin II.* Am J Physiol Heart Circ Physiol, 2000. **279**(2): p. H672-8.
76. Durante, W., et al., *Nitric oxide induces heme oxygenase-1 gene expression and carbon monoxide production in vascular smooth muscle cells.* Circ Res, 1997. **80**(4): p. 557-64.
77. Immenschuh, S., et al., *The rat heme oxygenase-1 gene is transcriptionally induced via the protein kinase A signaling pathway in rat hepatocyte cultures.* Mol Pharmacol, 1998. **53**(3): p. 483-91.
78. Tosaki, A. and D.K. Das, *The role of heme oxygenase signaling in various disorders.* Mol Cell Biochem, 2002. **232**(1-2): p. 149-57.
79. Alam, J. and J.L. Cook, *Transcriptional regulation of the heme oxygenase-1 gene via the stress response element pathway.* Curr Pharm Des, 2003. **9**(30): p. 2499-511.
80. De Keulenaer, G.W., et al., *Oscillatory and steady laminar shear stress differentially affect human endothelial redox state: role of a superoxide-producing NADH oxidase.* Circ Res, 1998. **82**(10): p. 1094-101.
81. Sogawa, K., et al., *Inhibition of hypoxia-inducible factor 1 activity by nitric oxide donors in hypoxia.* Proc Natl Acad Sci U S A, 1998. **95**(13): p. 7368-73.
82. Brune, B., A. von Knethen, and K.B. Sandau, *Nitric oxide (NO): an effector of apoptosis.* Cell Death Differ, 1999. **6**(10): p. 969-75.
83. Murphy, M.E. and H. Sies, *Reversible conversion of nitroxyl anion to nitric oxide by superoxide dismutase.* Proc Natl Acad Sci U S A, 1991. **88**(23): p. 10860-4.
84. Yano, S., et al., *Mechanical stretching in vitro regulates signal transduction pathways and cellular proliferation in human epidermal keratinocytes.* J Invest Dermatol, 2004. **122**(3): p. 783-90.
85. Zheng, J.L., C. Helbig, and W.Q. Gao, *Induction of cell proliferation by fibroblast and insulin-like growth factors in pure rat inner ear epithelial cell cultures.* J Neurosci, 1997. **17**(1): p. 216-26.
86. Scott-Burden, T., et al., *Platelet-derived growth factor suppresses and fibroblast growth factor enhances cytokine-induced production of nitric oxide by cultured smooth muscle cells. Effects on cell proliferation.* Circ Res, 1992. **71**(5): p. 1088-100.
87. Affymetrix, http://www.affymetrix.com.
88. Hancock, G.E., G. Kaplan, and Z.A. Cohn, *Keratinocyte growth regulation by the products of immune cells.* J Exp Med, 1988. **168**(4): p. 1395-402.
89. Lehninger, A.L., D.L. Nelson, and M.M. Cox, *Lehninger principles of biochemistry.* 4th ed. 2005, New York: W.H. Freeman. 1 v. (various pagings).
90. Liu, M., et al., *Stimulation of fetal rat lung cell proliferation in vitro by mechanical stretch.* Am J Physiol, 1992. **263**(3 Pt 1): p. L376-83.
91. Brickson, S., et al., *Oxidant production and immune response after stretch injury in skeletal muscle.* Med Sci Sports Exerc, 2001. **33**(12): p. 2010-5.
92. Capolicchio, G., et al., *Extracellular matrix gene responses in a novel ex vivo model of bladder stretch injury.* J Urol, 2001. **165**(6 Pt 2): p. 2235-40.

93. Otterbein, L.E., et al., *Exogenous administration of heme oxygenase-1 by gene transfer provides protection against hyperoxia-induced lung injury.* J Clin Invest, 1999. **103**(7): p. 1047-54.
94. Galbraith, R., *Heme oxygenase: who needs it?* Proc Soc Exp Biol Med, 1999. **222**(3): p. 299-305.
95. Berrar, D.P., W. Dubitzky, and M. Granzow, *A practical approach to microarray data analysis.* 2003, Boston, MA: Kluwer Academic Publishers. xv, 368.