CANCER CELL DYNAMICS

Terrence P McGarty
White Paper No 108
January 2014

We have expanded and combined work done previously into a single White Paper discussing cancer cell dynamics. We demonstrate the various ways to model cell growth, diffusion and flow across the body and integrate the genetic changes anticipated.

Notice

This document represents the personal opinion of the author and is not meant to be in any way the offering of medical advice or otherwise. It represents solely an analysis by the author of certain data which is generally available. The author furthermore makes no representations that the data available in the referenced papers is free from error. The Author also does not represent in any manner or fashion that the documents and information contained herein can be used other than for expressing the opinions of the Author. Any use made and actions resulting directly or otherwise from any of the documents, information, analyses, or data or otherwise is the sole responsibility of the user and The Author expressly takes no liability for any direct or indirect losses, harm, damage or otherwise resulting from the use or reliance upon any of the Author's opinions as herein expressed. There is no representation by The Author, express or otherwise, that the materials contained herein are investment advice, business advice, legal advice, medical advice or in any way should be relied upon by anyone for any purpose. The Author does not provide any financial, investment, medical, legal or similar advice in this document or in its publications on any related Internet sites.

Contents

1 CANCER MODELS FOR UNDERSTANDING, PREDICTION, AND CONTROL

Cancer cell dynamics is the study of how cancer cells mutate and move throughout the body. There has been a great deal of modelling of this phenomenon and in this paper we will provide some alternative approaches. We first will consider the single cell behavior and then second we will examine the ensemble or average tempero-spatial behavior. The intent is to develop a model depicting the average number of cancer cells of a specific type at a specific location at a specific time assuming we understand the mutational characteristics of the cell and the local behavior of cells.

We will now consider; what are the essential elements for modeling cancers? The first step is to re-establish the goals of a model and then its structure. Finally we will lead into the interrelationship between a model and the data which is used to justify it.

Many authors have developed models concerning pathways and also cancer. The books by Klipp et al and that of Szlassi et al are excellent overviews of the area with significant detail. The Klipp et al book is a truly superb discussion regarding pathways and modeling alternatives. The books by Bellomo et al and Wang are directed specifically at cancer modeling but unfortunately they lack adequate pathway dynamics to be of substantial use. Yet they are the only books available within the focused area.

One of the major issues of concern that drove this research is the explosion of "gene profiles" for various cancers. We have argued again and again that just because a gene is expressed differently in many cancers does not mean they are causative. They may just be correlated or confounded. The problem with the micro Array technology is it enables too many people to examine too many cancers and prognosticate on what may just be random events. We need verifiable and predictive models.

1.1 CORE MODEL

At the core, we want a model which reflects the following qualities:

1. Based Upon Reality: The model must at its core be based upon the known reality. It must conform to those factors which we currently know and understand. Namely it must reflect in its core the elements which we consider critical and the temporal and spatial dynamics of those elements. The model must be based upon a tempero-spatial system of measurable quantities; linked in some kinetic manner using reasonably well understood processes.

2. Predictability: Any modeling must, if it is to have any credibility, have the ability to predict, to say what will happen, and then to have that prediction validated. Although the ability may be statistical in nature the statistical confidence must be justifiable. We know all too well that many things are correlated, yet not causal, and not predictable.

3. Measurable: One must be able to measure and then predict the quantities which make up the model. Many of the modeling systems include proteins but they react in some zero-one format.

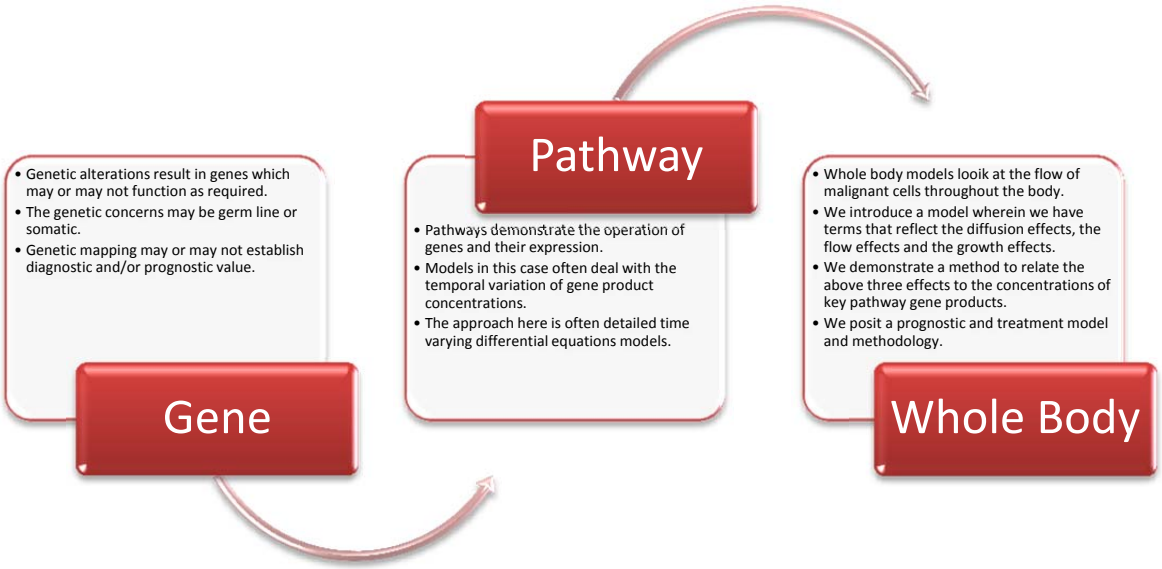
We know in reality that we have concentrations, or better yet specific numbers of proteins, produced in a cell. Yet we cannot yet measure the number of each of these proteins. We all too often can at best measure their presence or absence. However, is it not the case that it is the excess or the low density of some set of proteins which shift reactions, and that reactions are often concentration dependent.

4. Modellable: We want a system which can be modeled. It must reflect the measurable quantities in space and time and the tempero-spatial dynamics of them, using techniques that we can then use for prediction and validation.

In this paper we examine and analyze several models of cancer. Specifically we look at intracellular, extracellular and full body models. We attempt to establish a linkage between all of them. Many researchers have looked at the gene level, the pathway level and the gross flow of cancer cell level, namely whole body. Connecting them has been complex to say the least.

But herein we look at the pathway level and a whole body level and demonstrate the nexus, physically, and from this we argue that one can construct both prognostic tools as well as methodologies to deal with metastasis.

The following graphic lays out the flow of development and its implications as we detail them herein.



This examination that we will pursue following in the following manner:

1. We examine the current understanding of cancer and attempt to determine what type of model is best used in our approach.

2. We then examine briefly the issue of understanding what cause the two principle changes: proliferation and movement. Cancer cells do their damage primarily by proliferating and moving. The also demonstrate a loss of functionality, namely melanoma cells no longer act as melanocytes. Malignant blood cells no longer carry oxygen or fight infections.

3. Using an understanding of the genetic makeup of defective cancer cells we then examine their temporal and then spatial character. We look at several current models and then revert to a classic system model used now for over half a century in the engineering domain. Our goal is to develop an engineering approach, not a cellular approach.

4. We then examine the genetic mutations and consider developing a model which provides average behavior. The average is time and space specific but lets us specify average cell density per cell type per unit time at all related spatial locations.

5. Finally we demonstrate how this model may be applied to two areas. First we examine the identification of cancer cell changes by determining specific growth and propagation constants as related to specific gene expression changes and second we relate this to ascertaining such statistics as survival rates and hazard functions.

6. Finally we pose the problem of observability and controllability, namely the ability of having a viable model which can predict and ultimately be controlled.

What are the goals or objectives of our model? Simply:

1. Predictability: If we use the model we should be able to predict with some reasonable and measurable accuracy what will happen is few know the current state of some malignancy. This is the sine qua non of any model. Moreover we must agree as to what we are to predict. In our approach we will be examining average tumor burden from a specific malignancy as a function of body coordinates, location, and time. Namely if we are given the genetic dynamics, the probabilistic changes that can occur in control pathways, then we should be able to predict the changes in cells and from that determine the changes in numbers at specific places in space and time, on the average.

2. Controllability: We seek a model which is not only predictable but controllable, namely we must be capable of ascertaining measures that can drive the system to some desired state. Namely if we desire a tumor burden load of some small number then we should have a model, if at all possible, which allows us to drive the system to that state with some identifiable control mechanism.

These are the two simple goals of the model building. First we should be able to say what will happen. Second, we should be able to come up with a means to drive the tumor burden to a lower state by understanding what the mechanism are and having means, albeit to be developed means, to effect that change.

1.2 WHY THIS APPROACH

The question one may ask is why do we select this approach? Are not there many better ways to examine cancer cells using SNPs, GWAS, microarrays and the like? The answer is that we are no examining the cell as a bench biologist but examining cancer from a system perspective. As a Physicist examines electrons down through the quantum level, the Electrical Engineer looks at current, albeit an amalgam of electrons, and considers the agglomerate and then builds upward, having some certainty that the aggregate upon which they build will hold firm. We are attempting here to attain a similar footing.

There is a recent paper by Prof. Dougherty from Texas A&M which bemoans the state of some parts of science in the current environment[1]. As Dougherty so clearly states:

…science concerns relations between measurable variables and it is these relations that constitute the subject matter of science, scientific knowledge ipso facto is mathematically constituted…

Let me give a couple of examples of how this applies.

First, let us look at the world of genomics which I have been discussing herein for a while. The introduction of the microarray has allowed an explosion of data that has then allowed scientists to putatively argue some relationship between genes and cancers. Namely they go about examining say 9,000 prostate cancer patients and using microarrays primed for say 500 genes they conclude that say some 50 of these gene are seen in prostate cancer. They then allege that there is some actionable clinical relationship between the presence of the gene and the cancer. There is no underlying system model identifying this, just a microarray demonstrating that "oftentimes" these genes are under or over expressed.

Second, let us look at the BRAF V600 melanoma cases. Here unlike the above we have a case where one knows the RAF pathway and that loss of control of certain elements of that pathway lead to gene instabilities and thus a malignant expression. Therefore one targets the mutated RAF gene, the BRAF V600, and it results in a suppression of the malignancy, for a while. Then we had squamous cell carcinomas, but since the full pathway was known, go down one step and there was MEK and controlling it controlled the sequella. In this case there was a model, a system, and by logically following the system one found what the next step should be.

The above are two examples of how "science" is being done today in the area of gene related results. The second example is a Dougherty like science, namely it connects data to an underlying model which is predictable, and by using that the cancer is controllable, at least until another instability results. The first model, data collecting, is not really science as we accept it today. It is more akin to 19th century Botany, at best, where one goes out and collects specimens of plants and then tries to sew together a quilt of understanding to explain nature.

[1] http://bsb.eurasipjournals.com/content/2013/1/15

What Dougherty is focusing on is the Why question. When I recall Medical School, one is taught "What and how". What disease is it and how do I treat it. In contrast Engineering is first Why and then How. There is a strong dissonance when an Engineer is studying Medicine. At least forty years ago. An Engineer all too often keeps asking why, what is the underlying set of basic scientific principles that explain the phenomenon and how can I express them in a manner in which they can be used on a predictable basis. Why would drive many a Medical Professor to apoplexy. Medicine was for a long while the transfer down of "facts" and not validatable principles. The old adage at graduation that fifty percent of what one had just learned in Medical School was now invalid was a bit of a joke but sadly it was also true.

But as we move to Genomics we sadly see this trait arise again. There is a tension between those who want to have basic repeatable principles to build upon and those who believe that collecting data is the sine qua non. Let me give an example of a recent experience. Prof Lander at MIT is teaching an EdX course on Biology. Now Lander is brilliant and his style of teaching is in many ways classic MIT. Namely he highlights the basic principles, and then the student works through the Problem Sets developing the details for themselves. So far, so good. His first two three fourths of the course was fantastic. Then I noticed a subtle change, a change that, unless you were prepared to recognize would have slipped through the cracks. He slowly started giving a mixture or core predictable principles and cook book recipes. For example, we know that we can denature DNA because the base pair bonds are Hydrogen bonds, relatively weak, and the backbone Phosphate bonds are strong because they are ionic. Thus by heating the molecule we break the Hydrogen bonds first and then before we break the ionic bonds we can do our complementary additions, thus PCR works well.

On the other hand as he progressed to a discussion of Knock Out genes there were a collection of "tricks" or cook book recipes that were used. Why, for example did one get the modified DNA into the denatured gene the way he said? Well it just happens. Well nothing just happens. Fortunately bench Biologists have developed many "tricks", like alchemists, and as a result they have become a bit too comfortable with this unexplained bevy of tools, albeit indispensable, but in the long run self-defeating.

As Dougherty states when he examines data mining as an example of the Biologist's flair for data at all costs[2]:

Data mining and Copernicus share a lack of experimental design; however, in contradistinction to data mining, Copernicus thought about unplanned data and changed the world, the key word being 'thought.' Copernicus was not an algorithm numerically crunching data until some stopping point, very often with no adequate theory of convergence or accuracy. Copernicus had a mind and ideas. William Barrett writes, 'The absence of an intelligent idea in the grasp of a problem cannot be redeemed by the elaborateness of the machinery one subsequently employs'.

Or as M. L. Bittner and I have asked, 'Does anyone really believe that data mining could produce the general theory of relativity'? Data mining represents a regression from the achievements of three and a half centuries of epistemological progress to a radical empiricism,

[2] http://bsb.eurasipjournals.com/content/2013/1/15

in regard to which Reichenbach writes, 'A mere report of relations observed in the past cannot be called knowledge. If knowledge is to reveal objective relations of physical objects, it must include reliable predictions. A radical empiricism, therefore, denies the possibility of knowledge'. A collection of measurements together with statements about the measurements is not scientific knowledge, unless those statements are tied to verifiable predictions concerning the phenomena to which the measurements pertain.

What is Dougherty getting at? Simply, to reiterate the first quote: Science demands a marriage between data and models, to be true science it must be predictable and predictable based upon an embodiment in an abstraction.

Let me now apply this to genomics. Consider prostate cancer. The question is complex but can be asked; what is the first set of steps that lead to prostate cancer? Let us examine what we know:

First, we know many of the pathways. We know that the AKT pathway is critical, we know that c-MYC is a critical control element, we know that PTEN is often mutated, and we know that AR (Androgen Receptors) ultimately get mutated and we have metastatic growth. We pathways, we have relationships; we can demonstrate causality and results. Thus a modicum of a basis in reality exists. If one would use this pathway model and then search using microarrays matched against the model one arguable could iterate to improved models and improved predictability. The data without the model is useless and the model without the data is unverifiable.

Second, we can ask what sets the process off. Are all the changes due to mutations or more likely due to epigenetic insults? Thus when we look at MDS for example, we are looking at a hypermethylated set of blood stem cells. Something hypermethylated them and we know that since they are hypermethylated that the gene expression is repressed and thus cell proliferation of immature cells is a result. In prostate cancer, is the control mechanism lost because of a mutation, methylation, both, and in what order? Having a model allows one to validate and then iterate along a consistent trajectory of reality.

What does Dougherty have to say here?

While ignorance of basic scientific method is a serious problem, it is necessary to probe further than simply methodological ignorance to get at the full depth of the educational problem. Science does not stand alone, disjoint from the rest of culture. Science takes place within the general human intellectual condition. Biology cannot be divorced from physics, nor can either be divorced from mathematics and philosophy. One's total intellectual repertoire affects the direction of inquiry: the richer one's knowledge, the more questions that can be asked. Schrodinger comments, 'A selection has been made on which the present structure of science is built. That selection must have been influenced by circumstances that are other than purely scientific'

The point I believe he is making is that in the new world of Genomics, it is necessary to have a foundation that exceeds just the Laboratory and its tricks. One must understand that no matter what we think that every time we look at a cell, at an organism, we are looking at a system, at

some stochastic dynamical process wherein things move forward, albeit randomly, but in a way controlled by principles. We must look at the world wherein data is used not as an end in itself but as an iterative process with our mathematical world view. Thus the tools needed to view this world are extensive yet available. Engineers are trained to use them daily. Perhaps Genomics will grow to appreciate their essential import.

1.3 BATTLE OF WORLD VIEWS

It seems clear that people view the world by an application of the tools that they bring to that process. At one extreme a purely religious individual may very well accord any and all actions to an all-powerful deity. In contrast a scientist may very well see causal reality at all turns. Others may view things as just happening and that we must learn to live with near chaos. Thus having a toolbox of capabilities oftentimes take control of how we see the world and in turn react to it.

A classic example comes to the fore. Consider statisticians. A statistician looks at the data as nothing more than a collection of numbers and then given these numbers tries to make statements and assign to those statements some measure of their statistical validity. Statisticians do not it appears ever have an understanding or even an interest in the underlying phenomenon they have data on. In fact the often in my opinion eschew any interest as some biasing factor to their ultimate analysis. The classic example is a Logit analysis where the statistician posits some linear model in some exponential form and then uses regression to ascertain the relationship between probabilities and underlying factors. This approach is despite the underlying physical phenomenon.

A second example is the explosive use of microarrays to posit relationships between gene expressions and prognostic measures for various cancers. The microarray allows the researcher to find hundreds of correlated genes which are under or over expressed and statistics allow these measures to attain levels of significance independent of any underlying physical reality. The results may be harmful in the long run.

Another approach has been seen recently in the treatment of CML and melanoma. Understanding the physical process, what in that process is defective, and targeting the defect, we have therapeutics which assists in mitigating the disease. This is the reality approach, the reality world view, which we shall take herein. The key to this world view is the underlying question of: why? Why does a white cell proliferate so rapidly? Why does a melanocyte lose its ability to stay fixed in the basal layer? Why does a luminal cell in the prostate proliferate and create pseudo glands? The why is the underlying control mechanism of the cell, the pathways and the protein expression? Rather than just looking at expressions as measures unto themselves, we look at expressions as output measures of internal processes, processes that was understand, can model, and ultimately can control.

2 CHARACTERISTICS OF CANCER

As research has discovered how a cell functions the aberrant functions have become somewhat clearer. Cancer in simple terms is just a bad cell; a cell that just does not behave the way it is supposed to. However there are many reasons for why this may happen. No single aberrant form of behavior can determine a cancer; it often requires a plethora of many small but significant changes.

For example, take a simple melanocyte. It has developed from the neural crest during human embryonic development and ends up at the basal layer of the skin, for the most part. It then can develop a growth pattern which means its reproduces itself into a large cluster of melanocytes and each produces melanosomes and they end up in the top layers of the epidermis and we may have a freckle or mole. However, one characteristic has remained intact, namely the cells stay where they should be, in the basal layer. Now if one of these melanocytes loses its ability to stay put and starts wandering say upwards to the top of the basal layer then we often call this melanoma in situ. Namely, proliferation may not be bad but wandering can be.

In this section we examine some of the general characteristics of cancers. Historical understanding has been generally insightful, even before we understood the genetic changes that induce cancers. The questions we will examine are those that relate to what goes wrong; how do we model that, and how do we determine what the modelling equations and constants are?

Ultimately we will be faced with a continually evolving paradigm. The model will most likely not be fixed in space and time, but as we understand more we can add and delete accordingly. The discussions herein are merely a few initial steps, along with the many others being taken by other investigators.

2.1 THE BASIC MODEL

Let us begin by reviewing the basic characteristics of cancer cells and cancer in toto. What are the key characteristics of cancer? As Weinberg notes in the figure below, normal cells have the following functions:

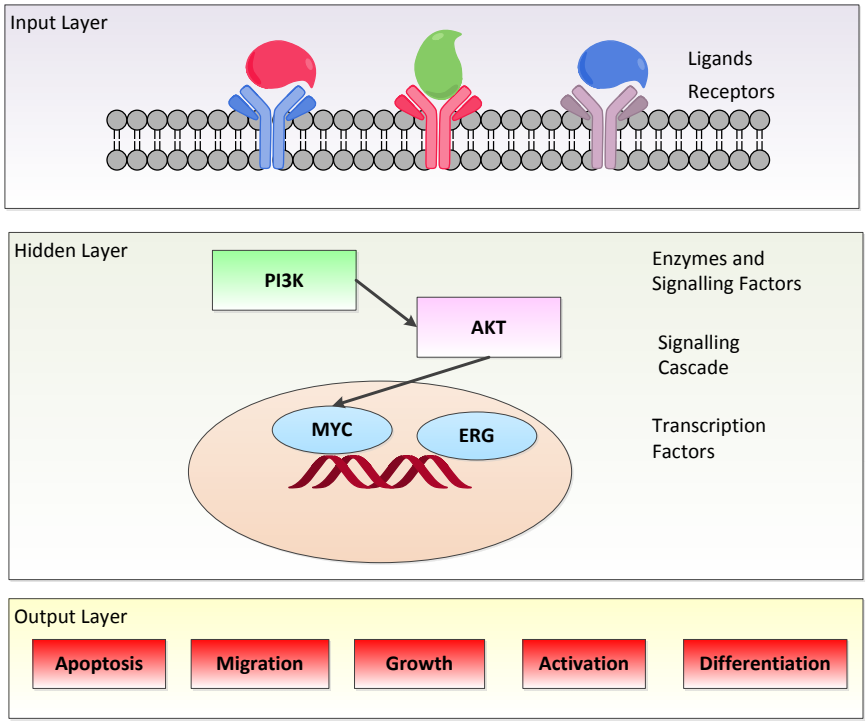
1. Growth: Growth is often distinguished from cell multiplication, or mitosis. Growth is cell expansion, the production of new proteins, the expansion of intra-cellular fluids, and the inclusion of exogenous factors found in normal cell life. In a sense cells consume and cells can get fat! The cells produce and absorb. They consume so they can both grow and duplicate, via simple mitosis.

2. Migration: Cells often have an ability to find their place and stay there. Blood cells flow out into the vascular system, and they tend to stay there until their natural course is complete. Basal cells remain basal as they progress up through the epithelium. Melanocytes remain melanocytes and remain in the basal layers, when benign.

3. Apoptosis: Cells die. They are programmed to do so and they do so in a clean and elegant fashion leaving no mess behind. This is in contrast to cell death with a diseased or damaged cell in which case the immune system must take over and remove the debris.

4. Differentiation: Cells have the ability to differentiate and perform specific functions.

5. Adhesion: Cells stick to adjacent cells, keeping their place in a collection of many other differentiated cells, and performing specific cellular functions.



The following represent several of the characteristics found in cancer cells.

1. Genetic Alterations: The genes in cancer cells are often markedly changed. We have demonstrated that the literature has for melanoma, and for many other cancers, laid out various genetic profiles, some connected with pathway alterations and others just recognized.

2. Cell Cycle Activation: Cells enter mitosis when activated generally y external growth stimuli and in turn activate growth pathways. It is possible that as a result of certain genetic pathway alterations that they may get activated endogenously, rather than exogenously. In either case we have activated the cell cycle and if not controlled the cell will complete mitosis and a doubling will occur. Unlimited cell cycle activation will result in the unlimited growth of cancer cells and the displacement of normal cells and the functions that those cells provide.

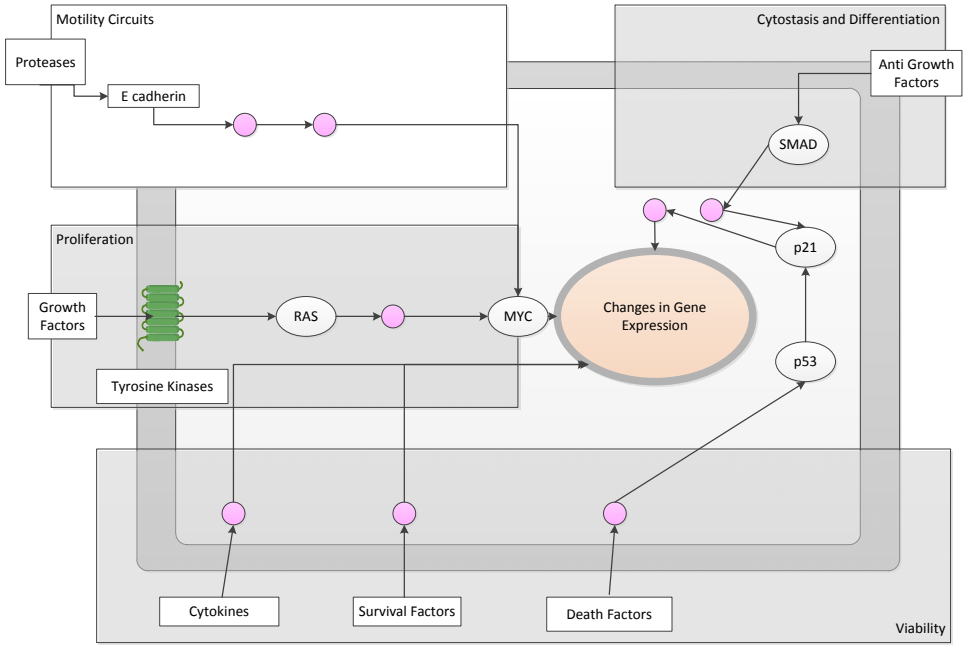
3. Loss of Normal Apoptosis: Loss of normal cell death is a major factor of cancer cells; they just do not die off.

4. Loss of Spatial Stability: The spatial instability is also a known attribute. The melanocyte for example just starts to wander, and then to replicate, and fails to die off through apoptosis.

5. Loss of Specific Functionality: Cancer cells lose all functionality. If a glandular cell, it no longer is. As a malignant melanocyte it no longer produces melanosomes for the skin.

6. Development of Alternative Metabolic Capabilities: This is in effect the Warburg Effect. Namely the metabolism of the cancer cell is anaerobic rather than oxidative.

7. Establishment of a Stem Cell Motif: We have discussed this at length before but it is simply the hypothesis that a small collection of cells control the metastatic process and that this small collection is a clone which may at this point be mutating independently. This implies that the remaining cancer-like cells are (i) not capable of separate metastatic, and (ii) if one removes the CSC then one halts the cancer and the remaining cells will go through some form of apoptosis. Also, (iii) the identification of the CSC is separate and apart from all other cells.



2.2 CANCER METABOLISM: AN EXAMPLE OF COMPLEXITY

When discussing the Total Cellular Dynamics model we often encounter the question of where does the cancer cell get its growth energy from? Does it seek paths which offer it the richest environment for expansion or are there other mechanisms that make this unnecessary? We briefly examine the Warburg Hypothesis. Warburg conjectured that cancer is caused by damage to respiration, oxidation, and the cells getting energy via fermentation.

As Warburg noted in 1956:

Cancer cells originate from normal body cells in two phases. The first phase is the irreversible injuring of respiration. Just as there are many remote causes of plague-heat, insects, rats-but only one common cause, the plague bacillus, there are a great many remote causes of cancer-tar, rays, arsenic, pressure, urethane- but there is only one common cause into which all other causes of cancer merge, the irreversible injuring of respiration.

The irreversible injuring of respiration is followed, as the second phase of cancer formation, by a long struggle for existence by the injured cells to maintain their structure, in which a part of the cells perish from lack of energy, while another part succeed in replacing the irretrievably lost respiration energy by fermentation energy. Because of the morphological inferiority of fermentation energy, the highly differentiated body cells are converted by this into undifferentiated cells that grow wildly-the cancer cells

Although the Hypothesis was and is consistent with observations the cause and effect are highly suspect. Yet it does pose the issue of energy infusion in cancer cells.

As Hsu and Sabatini remark concerning Warburg:

It is hard to begin a discussion of cancer cell metabolism without first mentioning Otto Warburg. A pioneer in the study of respiration, Warburg made a striking discovery in the 1920s. He found that, even in the presence of ample oxygen, cancer cells prefer to metabolize glucose by glycolysis, a seeming paradox as glycolysis, when compared to oxidative phosphorylation, is a less efficient pathway for producing ATP (Warburg, 1956). The Warburg effect has since been demonstrated in different types of tumors and the concomitant increase in glucose uptake has been exploited clinically for the detection of tumors by fluoro-deoxy-glucose positron emission tomography (FDG-PET).

Although aerobic glycolysis has now been generally accepted as a metabolic hallmark of cancer, its causal relationship with cancer progression is still unclear. In this Essay, we discuss the possible drivers, advantages, and potential liabilities of the altered metabolism of cancer cells. Although our emphasis on the Warburg effect reflects the focus of the field, we would also like to encourage a broader approach to the study of cancer metabolism that takes into account the contributions of all interconnected small molecule pathways of the cell.

And Hsu and Sabatini also demonstrate that this pathway control also controls apoptosis:

In addition to involvement in proliferation, altered metabolism may promote another cancer-essential function: the avoidance of apoptosis. Loss of the p53 target TIGAR sensitizes cancer cells to apoptosis, most likely by causing an increase in reactive oxygen species (Bensaad et al., 2006). On the other hand, overexpression of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) prevents caspase-independent cell death, presumably by stimulating glycolysis, increasing cellular ATP levels, and promoting autophagy (Colell et al., 2007). Whether or not GAPDH plays a physiological role in the regulation of cell death remains to be determined.

As Vander Heiden and Sabatini state:

In principle, the metabolic dependencies of cancer cells can be exploited for cancer treatment. For instance, a large fraction of human cancer is dependent on aberrant signaling through the PI3K/Akt pathway, and agents that target PI3K and various downstream signaling molecules are now in clinical trials.

The growing evidence that activation of PI3K causes increased dependency on glycolysis suggests that these agents may exert some of their effect by disrupting glucose metabolism. Drugs targeting key metabolic control points important for aerobic glycolysis, such as PK-M2 or LDH-A, might also warrant investigation as potential cancer therapies. In addition, the drugs developed to target metabolic diseases such as Type 2 Diabetes may have use in treating cancer. A number of retrospective clinical studies have found that the widely used diabetes drug metformin may offer a possible benefit in cancer prevention as well as improved outcomes when used with other cancer therapies.

Metformin and the more potent related compound Phenformin activate AMPK in cells, suggesting that Phenformin or other activators of AMPK might also be used as an adjunct to cancer therapy. Optimal use of these drugs will require a better understanding of cancer cell metabolism and identification of the signaling pathways that represent an Achilles' heel for cell proliferation or survival.

Metabolic tissues in mammals transform ingested food into a near-constant supply of glucose, glutamine, and lipids to balance the metabolic needs of both differentiated and proliferating tissues. Alterations in the appropriate balance of fuels and/or signal transduction pathways that deal with nutrient utilization may underlie the cancer predisposition associated with metabolic diseases such as diabetes and obesity.

A better understanding of how whole-body metabolism interacts with tumor metabolism may better define these risks and identify potential points of therapeutic intervention. In addition, it is possible that the cachexia associated with many cancers is exacerbated by the excess nutrient consumption by the tumor, which would affect whole-body metabolic regulation.

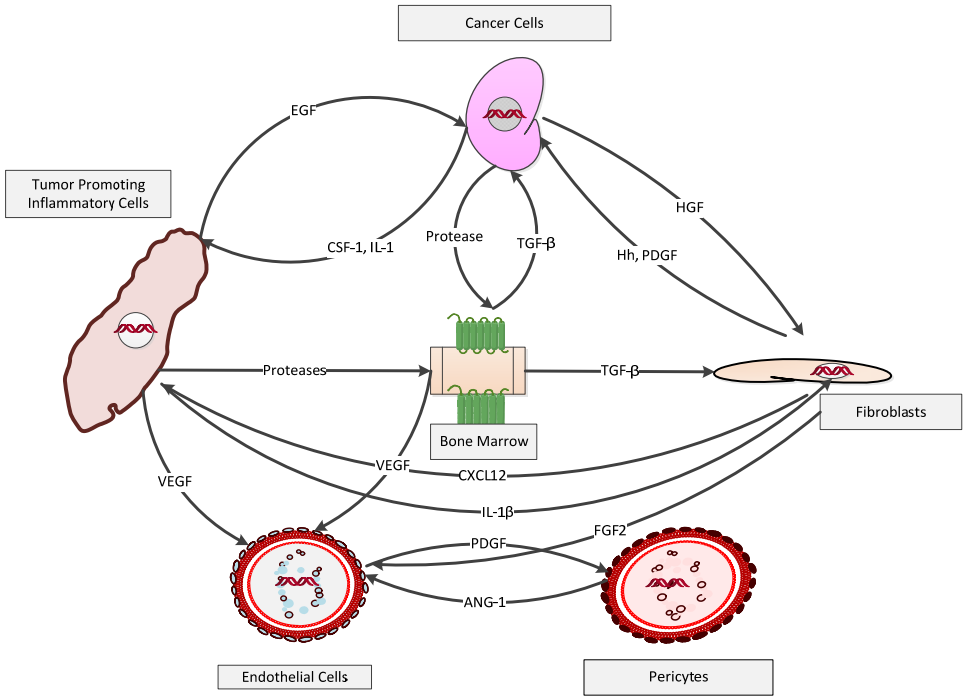
To this end, the potential role of dietary supplements and tight glucose control as adjuncts to cancer treatment is an active field of investigation.

The last comment has significant merit. The issue of glucose control, the impact of insulin related ligands and the pathway activation, all demonstrate this effect. One may then conjecture that when cancer cells migrate, they may do so along lines of higher glucose concentration. Thus, there is a preference for hematological migration. Again the Warburg Hypothesis is less of interest in the classic sense than in the ability to better understand migratory and diffusional behavior.

2.3 CANCER: AN EVOLVING PUZZLE

Cancer is an evolving puzzle. What we present here is not expected to be definitive. The challenge is that as we learn more about the cell and about cells in environments, it will most

likely be reflected in our understanding of cancer cell dynamics. For example, below we have a modification from Weinberg showing the interaction between cancer cells and other cells in the body.



Several recent papers have been published on the details of cancer genetics which make the understanding a continuous process of complexity. Let me first provide a brief précis of how we have progressed to this point:

1. The clone. It has been asserted that almost all cancers begin with a single aberrant cell, the clonal source. From this one cell we have generate everything else. One single cell then replicates in an uncontrolled manner.

2. The Vogelstein Paradigm: The Vogelstein Paradigm (VP) states that the clone is created in some predictable sequence of gene changes and that these changes can be detected and perhaps blocked.

3. The genetic profile: This concept uses the wealth, also excess, of gene mutation data available from microarray analysis to determine “profiles” for various cancers attempting to gain prognostic information as well as “individual” profiling for treatment. In many ways the micro array tool provides “too much data”, akin to the comment in Amadeus when the Emperor was asked about Mozart’s music, and he remarked “too many notes”. Namely the wealth of data is essential but the ability of the human processor is not quite up to it yet.

4. The pathway model: In this case we use pathways as a means to understand what is going wrong in a cell by cell basis. Then we try to block aberrant pathways to have the tumor no longer function as it has to that point. We have argued that this approach has a strong core, namely a model which can be verified and improved, but at the same time it lacks two major factors; (i) is does not deal with intercellular communications well enough, (ii) it does not deal with the issues of what causes the loss of gene activity and homeostasis well enough.

Now there have been several papers in NEJM discussing results on several cancers, kidney and AML, acute myeloid leukemia. Combined they tell and interesting tale. I have already commented on the kidney paper by Gerlinger et al but will add to it in this analysis.

As Gerlinger et al state:

Multiregion genetic analysis of four consecutive tumors provided evidence of intratumor heterogeneity in every tumor, with spatially separated heterogeneous somatic mutations and chromosomal imbalances leading to phenotypic intratumor diversity (activating mutation in MTOR*) and uniformity (loss-of-function mutation in* SETD2 *and* PTEN*).*

Of all somatic mutations found on multiregion sequencing, 63 to 69% were heterogeneous and thus not detectable in every sequenced region. Heterogeneous patterns of allelic imbalance were found in all tumors, and ploidy heterogeneity was found in two tumors. Therefore, we found that a single tumor-biopsy specimen reveals a minority of genetic aberrations (including mutations, allelic imbalance, and ploidy) that are present in an entire tumor.

Thus with this study we see significant genetic variability. The sequencing of genetic changes and the expectation of clonal consistency seems to be at variance.

In contrast, to justify the clonal progression, as Walter et al state regarding AML:

A unique aspect of the biology of leukemia is that hematopoietic cells freely mix and recirculate between the peripheral blood and the bone marrow. Clones that persist and grow over time must retain the capacity for self-renewal. Mutations in new clones must confer a growth advantage for them to successfully compete with ancestral clones. The result is that these secondary-AML samples are not monoclonal but are instead a mosaic of several genomes with unique sets of mutations; this mosaic is shaped by the acquisition of serial mutations and clonal diversification. Similarly, recent analysis of de novo AML samples with the use of whole-genome sequencing showed that relapse after chemotherapy is associated with clonal evolution and acquisition of new mutations.

Analysis of individual cancer cells may reveal additional layers of genetic complexity. Recent studies of B-cell acute lymphoblastic leukemia have shown that serial acquisition of cytogenetic abnormalities in that disease most often occurs through a branching hierarchy and only rarely follows a simple linear path…. Our study has several clinical implications.

First, the distinction between the myelodysplastic syndromes and secondary AML currently relies on manual enumeration of bone marrow myeloblasts, a standard that is subject to

interobserver bias but nonetheless drives major decisions about treatment for patients with small differences in myeloblast counts. Ultimately, identifying the patterns of pathogenic mutations and their clonality in bone marrow samples from patients with myelodysplastic syndromes should lead to greater diagnostic certainty and improved prognostic algorithms.

Neither studies presented intracellular pathways models which could be verified as state machines leading to malignant processes nor did they provide any basis for the genetic variations observed. These two factors will be essential in a better understanding of these diseases. However we see strong hematopoietic clonality and non-hematopoietic non-clonality.

The question one may ask is: do the cancer cells as they progress in a metastatic manner do so in a random ever changing manner unconnected from one another or is there some rational underlying physical basis for the changes in a manner in which the cancer has become an alter-organism in the human host? Is cancer a “slime mold” atop the human?

3 WHAT DO WE MODEL IN CANCER?

The key question we ask is; just what is it that we are modeling in cancer cell dynamics? We can start by saying that in a normal human we have a collection of many types of cells, all in some form of homeostasis. Each cell is doing what it is supposed to be doing. Some cells are being produced and some are dying off. Bone marrow is producing blood cells on a continuous basis whereas the brain has reached a point where the cells just function and do not reproduce. The extremes are many in the types of cells we deal with. Thus when one desires to find a model for cancer does one look at the body as a whose and assume that any cell can turn malignant or do we focus say on one type of malignancy and the follow its progress.

In this examination we shall follow the latter course. We will assume that we will follow a specific malignancy, without saying just which one, at this stage. For example we may assume we are to follow a melanoma or a prostate cancer development. This focus on a specific malignancy is but one way to develop a model.

Let us consider some options as to what "cells" we would want to model and measure.

1. Types of Cells: Typically we model a specific type of cell for a specific type of cancer. Thus we can examine basal or luminal cells for prostate cancer and use know genetic factors.

2. Types of Gene Expression: By focusing on specific cells and specific cancers we than have the ability to focus on specific genetic pathways. Here we assume that we understand how cell instability in a cancer environment is developed and managed via genetic pathways. Also since we focus on genetic pathway effects we believe that we can handle epigenetic factors, in effect considering epigenetic blocking or loss thereof as a pathway promoter or repressor effect.

3. Number of Calls at a Specific Location: Our focus will be the counting of specific cells at specific times of a specific genetic makeup. Thus in a benign environment we may have a baseline set of basal or luminal cells in a prostate gland. The genetic makeup of those cells will also be baseline. Now as we induce changes in the genetic makeup we see proliferation and loss of local stability. These factors will be phenotypic effects of changes in genetic expressions. We start with the genetic expression and then proceed to what that does to cell proliferation, namely number of cells, and cell movement, that is number of cells at specific locations.

4. Total Number of Specific Types of Cells: We can develop equations for actual cell counts but such a development is but one sample path of a stochastic process. Thus our final approach will be to develop ensemble averages or specific times and locations. These averages will reflect a reasonable ensemble behavior of the cancer being analyzed and will be reflective of the disease.

3.1 INTRACELLULAR GENE DYNAMICS:

This type of model focuses on the genes, and their behavior. It is basically one where we examine the gene type and its product.

3.2 INTRACELLULAR PROTEIN DYNAMICS:

This type of model falls in several subclasses. All begin with protein pathways and the "dynamics" of such pathways. But we have two major subclasses; protein measures and temporal measures. By the former we mean that we can look at the proteins as being on or off, there or not there, or at the other extreme looking at the total number of proteins of a specific type generated and present at a specific time. By the latter, namely the temporal state, we can look at the proteins in some static sense, namely there or not there at some average snapshot instance, or we can look at the details over time, the detailed dynamics. In all cases we look at the intracellular dynamics only.

Let us consider the two approaches.

i. On-Off: In this approach the intracellular relationships are depicted as activators or inhibitors, namely if present they allow or block an element in a pathway. PTEN is a typical example, if present it blocks Akt, if absent it allows Akt to proceed and enter mitosis. p53 is another example for if present we have apoptosis and if absent we fail to have apoptosis. These are simplistic views. This is a highly simplistic view but it does align with the understanding available say with limited microarray techniques. This is an example of the data collection defining what the model is or should be.

ii. Density: This is a more complex model and it does reflect what we would see as reality. The underlying assumptions here are:

a. Genes are continually producing proteins via transcription and translation.

b. Transcription and translation are affected at most by proteins from other genes acting as repressors or activators. There are no other elements affecting the process of transcription and translation. Not that this precludes any miRNA, methylation, or other secondary factors. We shall consider them later. In fact they may often be the controlling factors.

c. The kinetics of protein production can be determined. Namely we know the rate at which transcription and translation occur in a normal cell or even in a variant. That is we know that the production rate of proteins can be given by:

$$\frac{\partial n_j(t)}{\partial t} = \sum_{k=1}^{K} \lambda_k n_k(t)$$

Here we have production rates dependent on the concentration of other proteins. The processes related to consumption are not totally understood (see Martinez-Vincente et al). We understand cell growth, as distinct from mitotic duplication, but the growth of a cell is merely the expansion of what was already in the cell when at the end of its mitotic creation. In contrast, we understand apoptosis, the total destruction of the cell, we also understand that certain proteins flow outside the cell or may be used as cell surface receptors, but the consumption of these is not fully understood. Yet we can postulate:

$$\frac{\partial n_j(t)}{\partial t} = -\sum_{k=1}^{K} \mu_k n_k(t)$$

This is based upon the work of Martinez-Vincente et al which states:

All intracellular proteins undergo continuous synthesis and degradation (Mortimore et al., 1989; Schimke, 1970). This constant protein turnover, among other functions, helps reduce, to a minimum, the time a particular protein is exposed to the hazardous cellular environment, and consequently, the probability of being damaged or altered. At a first sight, this constant renewal of cellular components before they lose functionality may appear a tremendous waste of cellular resources.

However, it is well justified considering the detrimental consequences that the accumulation of damaged intracellular components has on cell function and survival (Goldberg, 2003). Furthermore, protein degradation rather than mere destruction is indeed a recycling process, as the constituent amino acids of the degraded protein are reutilized for the synthesis of new proteins (Mortimore et al., 1989; Schimke, 1970).

The rates at which different proteins are synthesized and degraded inside cells are different and can change in response to different stimuli or under different conditions. This balance between protein synthesis and degradation also allows cells to rapidly modify intracellular levels of proteins to adapt to changes in the extracellular environment. Proper protein degradation is also essential for cell survival under conditions resulting in extensive cellular damage. In fact, activation of the intracellular proteolytic systems occurs frequently as part of the cellular response to stress (recently reviewed in Cuervo, 2004b; Goldberg, 2003). In this role as 'quality control' systems, the proteolytic systems are assisted by molecular chaperones, which ultimately determine the fate of the damaged/unfolded protein (Fig. 1).

Damaged proteins are first recognized by molecular chaperones, which facilitate protein refolding/repairing. If the damage is too extensive, or under conditions unfavorable for protein repair, damaged proteins are targeted for degradation. Protein degradation is also essential during major cellular remodeling (i.e. embryogenesis, morphogenesis, cell differentiation), and as a defensive mechanism against harmful agents and pathogens (recently reviewed in Cuervo, 2004a; Klionsky, 2005).

We have also discussed this process with regards to the function of ubiquitin, which marks proteins for elimination. As Goldberg states:

Proteins within cells are continually being degraded to amino acids and replaced by newly synthesized proteins. This process is highly selective and precisely regulated, and individual proteins are destroyed at widely different rates, with half-lives ranging from several minutes to many days. In eukaryotic cells, most proteins destined for degradation are labelled first by ubiquitin in an energy requiring process and then digested to small peptides by the large proteolytic complex, the 26S proteasome.

Indicative of the complexity and importance of this system is the large number of gene products (perhaps a thousand) that function in the degradation of different proteins in mammalian cells. In the past decade, there has been an explosion of interest in the ubiquitin–proteasome pathway, due largely to the general recognition of its importance in the regulation of cell division, gene expression and other key processes[1]. However, the cell's degradative machinery must have evolved initially to serve a more fundamental homeostatic function — to serve as a quality-control system that rapidly eliminates misfolded or damaged proteins whose accumulation would interfere with normal cell function and viability.

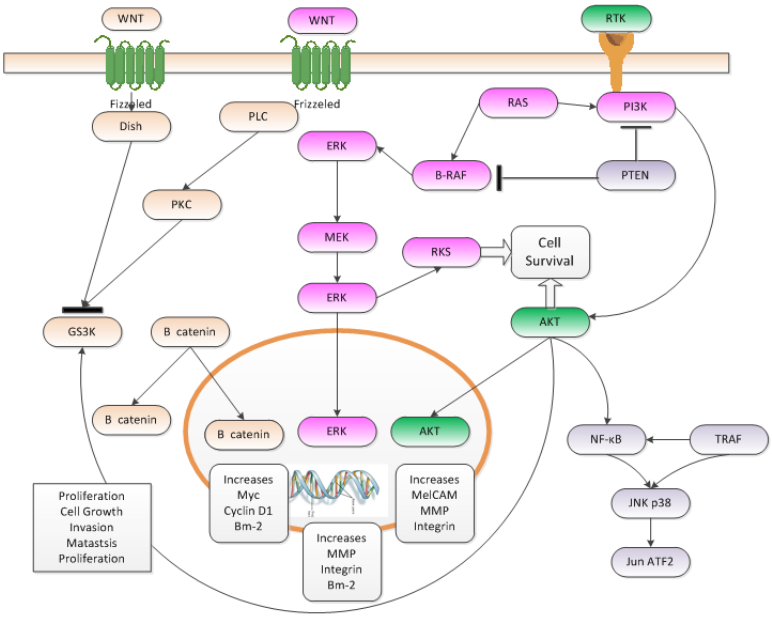
Also we refer to the recent review work of Ciechanover which details the evolution of this understanding.

In contrast the proteins are consumed and thus the negative sign. In toto we have:

$$\frac{\partial n_j(t)}{\partial t} = \sum_{k=1}^{K} (\lambda_k - \mu_k) n_k(t)$$

as a total balance of proteins. This assumes we have a production mechanism for each of the proteins, namely their genes and the activators and repressors as required.

d. Pathway Dynamics must be meaningful. Let us consider the pathway as shown below. This is a typical melanoma pathway we have shown before.



Now let us consider PTEN blocking BRAF and Akt. Now physically it is one molecule of PTEN needed for each molecule of BRAF and PI3K. But what if we have the following:

$$n_{PTEN}(t) < n_{BRAF}(t)$$
$$n_{PTEN}(t) > n_{PI3K}(t)$$

Here we have PTEN blocking some but not all the BRAF and PTEN blocking all the PI3K. At least at time t. Do we have an internal mechanism which then produces even more PTEN? One must see here that we are looking at the actual numbers of PTEN, real numbers reflecting the production and destruction rates. We know for example that if we have a mutated BRAF then no matter how much PTEN we have an unregulated pathway.

Now it is also important to note that this "model" and approach is distinct in ways from classic kinetics, since the classic model assume a large volume and concentrations in determining kinetic reaction rates of catalytic processes. Here we assume a protein binds one on one with another protein to facilitate a pathway.

Thus knowing the dynamics of individual proteins, and knowing the pathways of the proteins, namely the temporary adhesion of a protein, we can determine several factors:

1. The number of free proteins by type
2. The pathways activated or blocked
3. The resultant cellular dynamics based on activated pathways.

It should be noted that we see pathways being turned on and off as we produce and destroy proteins. There is a dynamic process ongoing and it all depends on what would be a stasis level of proteins by type. The question is; are cells in stasis or are they in a continual mode of regaining a temporary stasis?

This also begs the question, that if as we have argued, that cancer is a loss of stasis due to pathway malfunction, then can this be a process of instability in the course of a normal cell? Namely is there in the dynamics of cell protein counts, unstable oscillator type modes resulting in uncontrolled mitotic behavior. Namely can a cell get locked into an unstable state and start reproducing itself in that state, namely an otherwise normal cell.

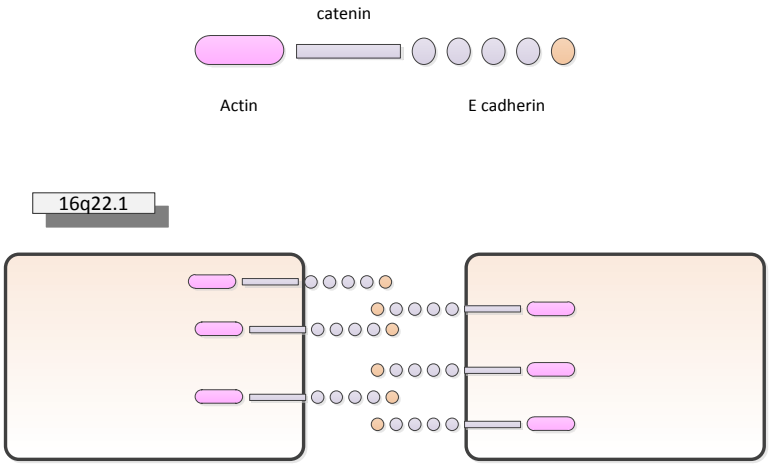
e. Total intracellular dynamics can be modeled yet the underlying processes are still not understood and the required measurements are yet to be determined.

3.3 INTERCELLULAR DYNAMICS:

Here we look at the intercellular dynamics as well, not just as a stand-alone model. By this methodology we look at intercellular communications by ligand binding and the resulting activation of the intracellular pathways. We must consider both the intercellular signalling between like cells but also between unlike, such a white cells perhaps as growth factor inhibitors and the like. We also then must consider the spatiodynamics, namely the "movement" of the cells or in effect the lack of fixedness or specificity of function. This becomes a quite complex problem.

There are two functions we examine here:

a. Intercellular binding or adhesion: E cadherin is one example that we see in melanocytes. Pathway breakdown may result in the malfunctioning of E cadherin.



The above demonstrated E cadherin in melanocyte-keratinocyte localization. The bonds are strong and this stabilizes the melanocyte in the basal layer. If however the E cadherin is compromised then the bond is broken, or materially weakened, and the melanocyte starts to wander. Movement for example above the bottom of the basal layer and upwards is pathognomonic of melanoma in situ. Wandering downward to the dermis becomes a melanoma. Thus the pathways activating E cadherin production is one pathway essential in the inter-cellular dynamics.

b. Ligand production and receptor production: Here we have cells producing ligands, proteins which venture out of the cell and become signalling elements in the intercellular world. We have the receptor production as well, where we have on the surface of cells, various receptors, also composed of cell generated proteins, which allow for binding sites of the ligands and result in pathway activation of some type. For example various Growth Factors, GF proteins, find their way to receptors, which in turn activate the pathways. Wnt is an example of one of these ligands which we have shown above.

It can also be argued that as ligands are produced and as the "flow" throughout the intercellular matrix, we can obtain effects similar to those in the Turing tessellation models. Namely a single ligand may be present everywhere but density of ligands may vary in a somewhat complex but determinable manner, namely is a wavelike fashion.

Thus, as with intracellular proteins, we have with intercellular ligands the following:

$$q(x,t) = E\left[Number\ of\ ligand\ Proteins; x, t\right]$$

and we would have some dynamic process as:

$$\frac{\partial q(x,t)}{\partial t} = R(x,t)\frac{\partial^2 q(x,t)}{\partial x^2} + \ldots$$

This is akin to the Turing model used in patterning of plants and animals. Namely the concentration of a ligand, and in turn its effect, may be controlled by

3.4 TOTAL CELLULAR DYNAMICS:

In this case we would want a model which reflects the total body spatiotemporal dynamics This type of models is an ideal which may or may not be achievable. In a simple sense it is akin to diffusion dynamics, viewing the cancer cells as one type of particle and the remaining body cells as another type. The cancer cells have intercellular characteristics specific to cancer and the body cells have functionally specific characteristics. Thus we could ask questions regarding the "diffusion" of cancer cells from a local point to distant points based upon the media in between. The "rate" of such diffusion could be dependent upon the local cells and their ability for example to nourish the cancer cells as well. In this model we could define an average concentration of cancer cells at some position x and time t as:

$$p(x,t) = E\left[Number\ of\ Cancer\ Cells; x,t\right]$$

and we would have some dynamic process as:

$$\frac{\partial p(x,t)}{\partial t} = D(x,t)\frac{\partial^2 p(x,t)}{\partial x^2} + \ldots$$

This is a diffusion like equation and is a whole body equation. Perhaps knowing what the rate of diffusion is on a cell by cell basis may allow one to determine the most likely diffusion path for the malignancy, and in turn direct treatment as well.

This is of course pure speculation since there has been to my knowledge any study in this area. Except one could imagine a system akin to PET scans and the like which would use as input the surface markers from a malignancy and then the body diffusion rates to plot out in space and time the most likely flow of malignant cells and thus plan out treatment strategies. Although this model is speculative we shall return again to it in a final review of such models since it does present a powerful alternative.

This concept of total cellular dynamics is in contradistinction to the intercellular transport. In the total cellular dynamics model we regard the model as one considering the flow of altered cells across an existing body of stable differentiated cells.

We may then ask what factors drive cancer cells to what locations? One may putatively state that cancer cells will follow the path of least resistance and/or will proceed along "flow lines" consistent with what propagation dynamics they may be influenced by.

4 INTRACELLULAR DYNAMICS

The question now posed is; how do we model the tempero-spatial dynamics of cancer cells? There has been a great deal of work in this area and we shall examine but a few examples. However we propose ultimately to use a system model. A system model is one where we determine the tempero-spatial distribution of cells and the fact that they are propagated by changes in their internal genetic operations. We do not worry about the specifics of what is happening inside the cell, albeit it does affect cell specific characteristics. We focus on gross cellular characteristics.

We now consider in more formalistic terms how to develop models. There are four approaches under consideration:

Boolean Networks: The model assumes discrete time on or off nets controlled by the dynamics of the pathway. The Boolean model consists of two elements: first, an assumption that a gene or protein is on or off, second, that the control of that constituent is a Boolean control consisting of AND, OR and NOT Boolean type functions dependent upon precursor constituents. We discuss this in some detail related to PTEN. The Boolean approach is interesting in that it works well at times identifying non-dynamic pathways using massive micro-array data.

Bayesian Networks: The Bayesian model assumes that the value may be analog not binary and that there exists a graph which is what is the desired output. It uses a Bayes model which yields the best fitting graph given data from a microarray and assuming that there are linkages consisting of probabilities in a directed manner. The Bayesian may become time varying as well. As with the Boolean approach it works well with ascertaining linkages and graphs using microarray data and in this case the microarray data may be of an analog nature not just binary.

Reaction Rate Models: Many researchers have tried to focus on what are called reaction rate models. This is chemical kinetic models as applied to a cell. They have proposed various models to account for cellular instability. We have examined many of them but after considerable analysis consider them to have substantial defects. Namely they assume high concentrations of substances. In a single cell there may be a small and countable number of proteins and thus this approach may be used inappropriately.

State Models: These are continuous time and continuous state models of specific pathway concentrations where we employ enzymatic reaction rate models that we have discussed previously. We have used these models in other circumstances and they have proven their worth. Also these models are often more amenable to stochastic effects as well as examination for observability and controllability.

4.1 BOOLEAN MODELS

The simplest model for pathways is the Boolean model. This has been described in Klipp et al in simple terms. A classification of Boolean Networks has been done in Gershenson. It is based upon three simple assumptions.

First, genes and their products are either on or off, expressed or not expressed. This denies concentration effects.

Thus as an example we may have PTEN and Akt, and they are either 1 or 0. Thus we have the four states, {PTEN, Akt} as 00, 10, 01, and 11. Not all may exist, we will show that next.

Second, the effects on a set of gene products on other gene products can be expressed by a Boolean expression.

Thus, for example we have the Binary or Boolean expression:

Akt = NOT PTEN.

Thus if PTEN is 0 then Akt is 1 and likewise if PTEN is 1 then Akt is 0. The states 11 and 00 are not allowed.

Third, this is a discrete time state system where we have the kth instance influencing the k+1 th state.

Thus as above we should state:

Akt (k+1) =NOT PTEN (k)

We may add asynchrony and probabilistic transitions as well to the model.

Now let us formalize the model a bit. We assume that there are N genes or gene products that are of interest. Then the state of the system at any time is:

$$x(k) = \begin{bmatrix} x_1(k) \\ \dots \\ x_N(k) \end{bmatrix}$$

where the xs can take on only 0 or 1 in value, on or off.

Now we assume that for each x (k) we have some Boolean equation of the following type:

$$x_j(k+1) = (x_1(k) AND x_3(k)) NOT x_4(k) OR x_6(k)$$

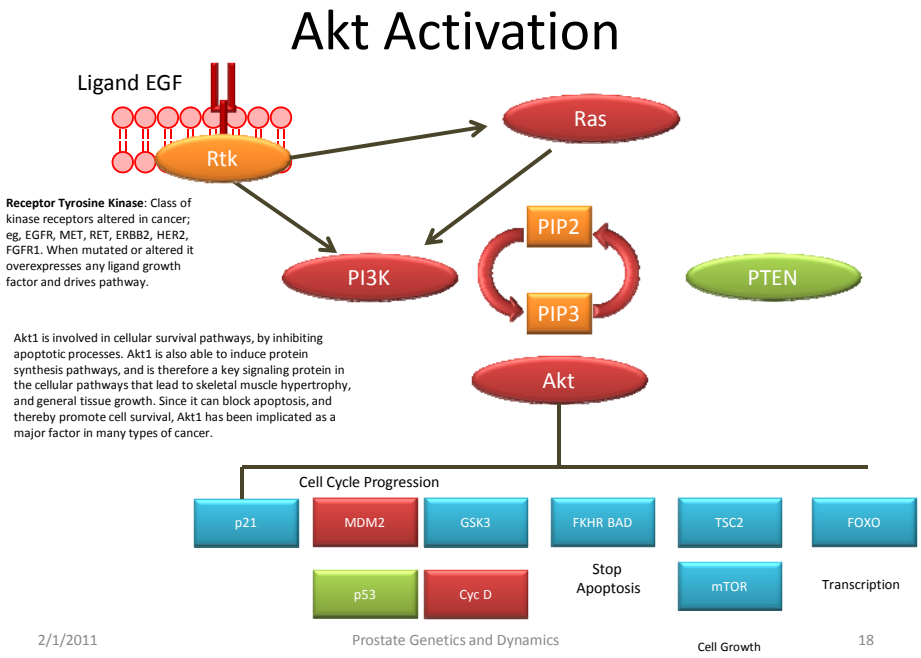
We now ask; where did this equation come from? We defer that for the moment. But let us assume that we can find that for every one of the x values. The we have:

$$x(k+1) = F(k+1,k)$$

where

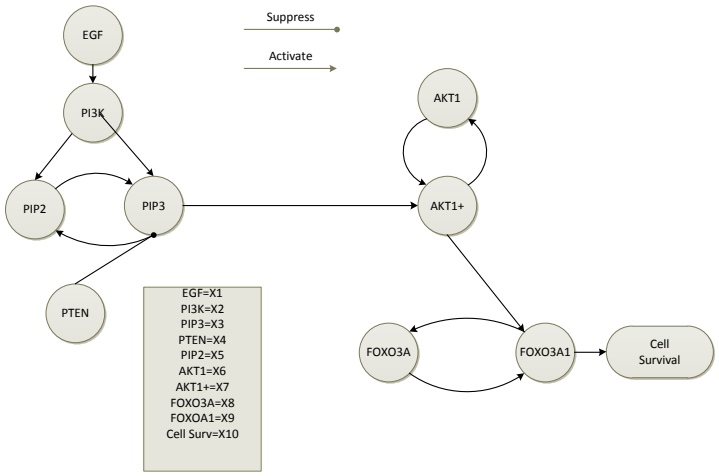
$$F(k+1,k) = \begin{bmatrix} f_1(k+1,k; x_1(k)...x_N(k)) \\ ... \\ f_N(k+1,k; x_1(k)...x_N(k)) \end{bmatrix}$$

Now let us return to our PCa model and look at a simplified set of it as shown below. First, from Lin et al we have for the simple model showing that "*PTEN Decreases AR Protein Levels via Promotion of AR Degradation*" which we show below;



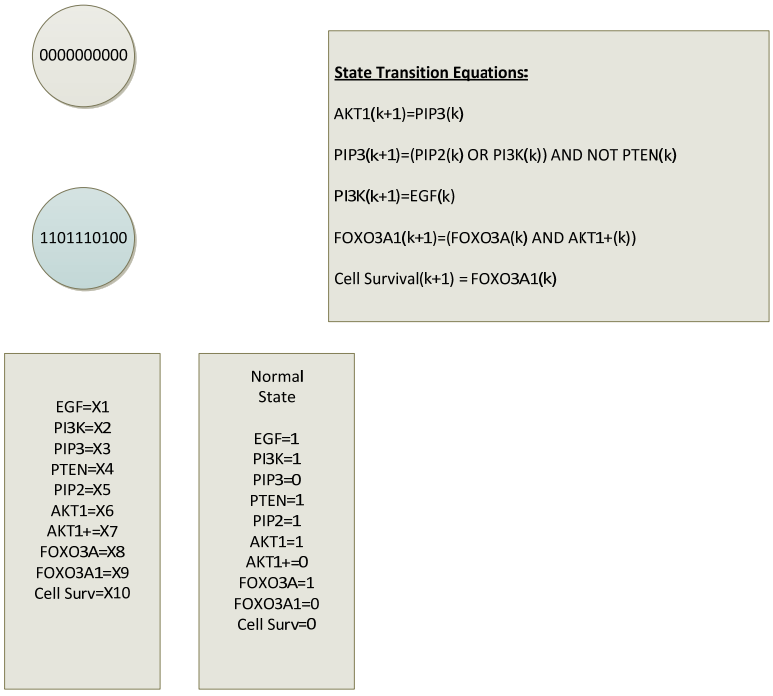
:

Then from the above we can generate a simple model of pathways with some detail as below:



Now in this model we assume that an EGF ligand drives PI3K which moves PIP3 to activate AKT to AKT+1 the activated form which drives FOXO to enhance cell survival. We see 10 states all of which are binary. Ideally we would have 2^{10} possible states, roughly a million. But the state transitions will prohibit this.

The following depict the transition equations, the Boolean f functions from above, as we have somewhat arbitrarily created them, and the stable state.



Note that we can assume that EGF is one and PTEN is 1 and thus PTEN blocks everything. Note also that only 5 state equations are specified. The others are just stable states. We can now take

any states, namely ones where EGF and PTEN go from 0 to 1 or 1 to 0 and see what happens. Note that we have in this model the assumption that PI3K needs EGF. Also if PTEN is on then it suppresses cell growth via FOXO.

One can use microarray technology to validate this model. However it has significant drawbacks.

First it is binary. Real world models are not that way.

Second it is discrete time, again not a reflection of reality.

Third, it assumes binary reactions, and the real reactions are a bit more subtle.

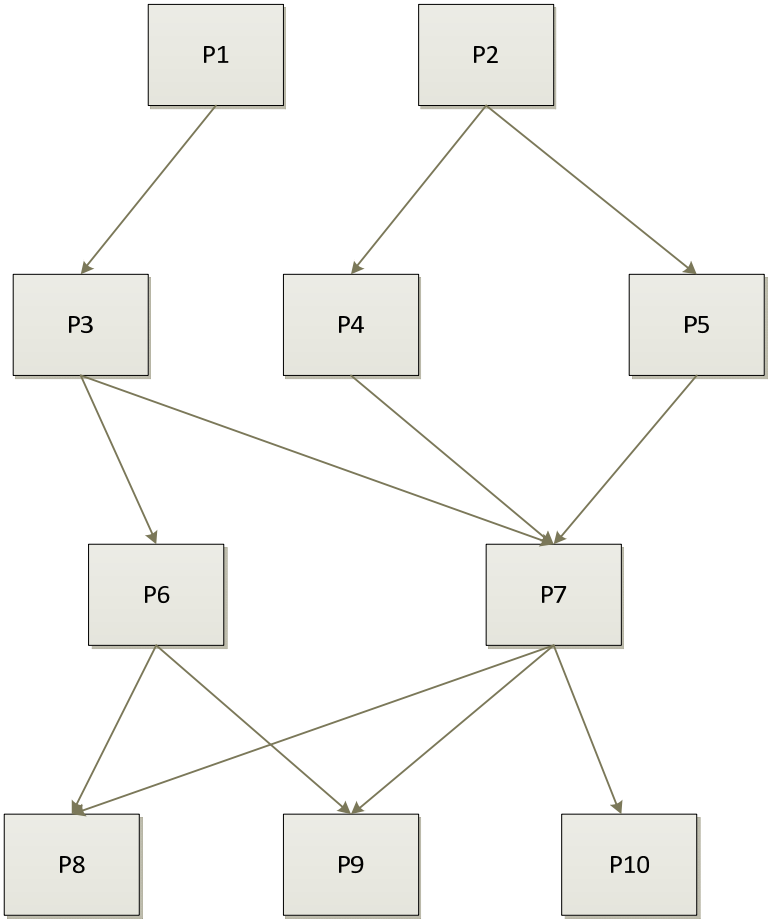
The use of the pathway data is oftentimes effective at the gross level (see Kim et al). Also the ability of correlating pathway models with microarray data is also of significant value. Oftentimes the microarray data itself is binary and thus it maps fairly well onto this model (see Driscoll and Gardner).

Expansion of this model to non-discrete times and random transitions has been shown in the wok of Shmulevich et al (three papers in 2002, 2002, 2002)

4.2 BAYESIAN MODELS

In the development of a Boolean Network we assumed two things; first that the proteins or other relative constituent was either on or off and second that the interaction between constituents was controlled via some Boolean function of AND, OR, NOT type combinations. Now in the Bayesian world we assume Bayes type relationships and then try to infer dependency thereupon.

Let us begin by looking at the following network. We assume that the Ps are influencing the Ps they are connected to in some manner. For simplicity we shall assume they are all positive influences.



Now in a Bayes world, we generally have some data and we then try to infer from the data the structure as we see it above.

Let us look at the above flow. Here we have assumed a priori that there are well defined influencing factors. But, what if all we had was10 products and we just thought that they somehow interact? What we want to do is use some logical method to arrive at the chart above. How could we approach this?

Let us assume we have some microarray data. Let us further assume that it is simple microarray data showing genes on or off.

We could write in general the following (see Theodoridis and Koutroumbas, pp 64-68):

$$p(P10,P9,P8,...P1) = p(P10|P9,...P1)p(P9|P8,...P1)...p(P2|P1)p(P1)$$

Now that does not reflect the graph, G, which we had shown at the top. One may look at the set of all possible graphs, say:

$$\Gamma = \{G_i\}$$

where we have any possible G as a set. We are really interested in the "best" G give say a data set D.

So let us go to the data set, D. What can we use to assist in determining the best possible G.? Let us assume we have a microarray. The microarray presents data in one of two forms; active or no effect. Namely the array elements lights up if active and does not light up is not active. Let us assume we have 20 samples and we test for the ten products. We get an array as below. Here red is active.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
P1																				
P2																				
P3																				
P4																				
P5																				
P6																				
P7																				
P8																				
P9																				
P10																				

From this we have a data set given by;

$$d_i = \begin{bmatrix} d_{i,1} \\ \dots \\ d_{i,10} \end{bmatrix}$$

and

$$D = \{d_1, \dots d_N : N = 1, \dots, 20\}$$

We can now look at the problem in a Bayesian sense. Let G be the graph and D the data as we demonstrated above. Now as a Bayes approach we can seek:

$$max\left[log\, P\left[G|D \right] \right] =$$

$$max\left[log\, P\left[D|G \right] + log\, P\left[G \right] + c \right]$$

We thus seek to maximize the Bayes score. We seek the G which given the D gives the best result.

Now we can model the system as follows:

$$P_i = \sum_{k=1,\neq i}^{10} a_{i,k} P_k$$

and

$$d_i = b_i P_i$$

The a values are determined by the graph G. We can add noise to the measurement and even to the system. We can then find the set of a that given the D set minimize the measure we stated above. Since we can assume that we have:

$$P_i = \sum_{k=1,\neq i}^{10} a_{i,k} P_k + n_i$$

and

$$d_i = b_i P_i + w_i$$

where they are zero mean Gaussian, then the metric used to attain a minimum on the data set D to determine the graph G is simple, yet a complex calculation.

Namely we have (see Shmulevich and Dougherty):

$$max\left[log\, P\left[G \middle| D \right] \right] =$$

$$max\left[log\, P\left[D \middle| G \right] + log\, P\left[G \right] + c \right]$$

or

$$max\left[-\sum_{k=1}^{10} \sum_{i=1}^{20} \left(d_{i,k} - \hat{d}_{i,k} \right)^2 \right] + log\, P\left[G \right]$$

where we have used the estimated d for the specific graph G being tested. We do this for every possible graph, G. That is a very computationally complex calculation. Indeed it can become non-calculateable.

We demonstrate in the next section that we can measure analog values for the measurements as well as for the desired connections.

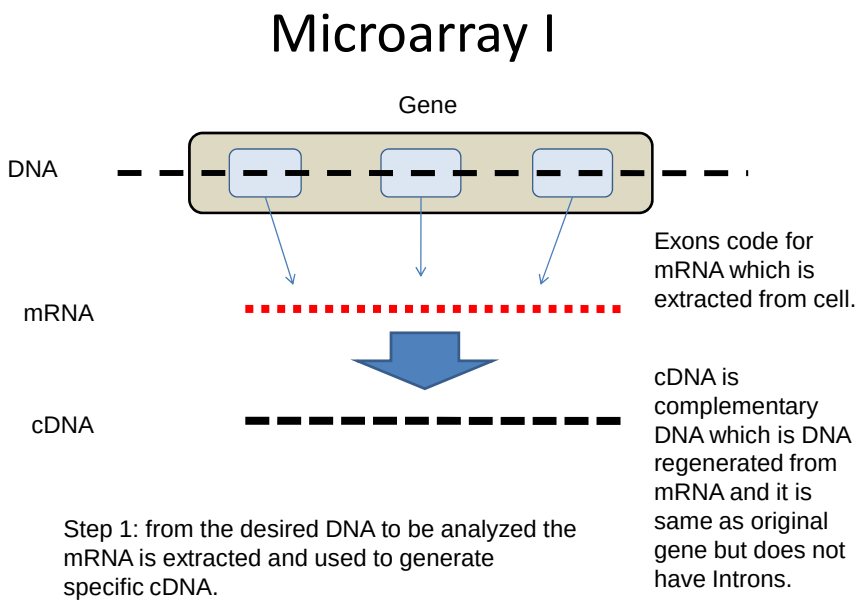
We briefly look at microarrays in some detail since we will use them in the Bayesian case as well in the full system modeling case.

Microarrays are simply a set of matrices where we take samples from many cells and place them in an N by M matrix, say using rows for genes and columns for cell samples. By selecting the rows such that we can place in each row a gene specific binding site, which we can even tag with some visible marker, we can then determine if on a sample by sample basis we have genes being expressed, namely that they bind at the binding site. It is also possible to attempt to ascertain the concentration of the gene or gene product by having the amount of binding in each sample be

reflected by a linear or similar color change. We can say have red for no binding, yellow for some and green for a great deal, or whatever color combination we so desire. Indeed it may then be possible to calibrate for relative concentrations in each cell entry via a colorimetric measurement.

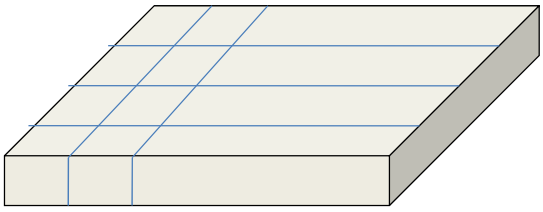
Microarrays is a unique approach which allows for the analysis of millions of samples, it is a marriage of high tech solid state chip technology with DNA bonding. We describe it in the following four steps, each step accompanied by a Figure.

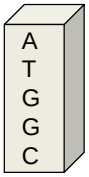
Step 1: The first step in a micro array is the production of cDNA, or complementary DNA. cDNA is that set of nucleotides which account for the encoding of mRNA. It does not include the non-coding regions which are the introns.



Step 2: In a separate environment we make the microcell. This is created in a manner identical to the making of integrated circuits which entails photo-masking techniques. Instead of silicon we used nucleotides. The array has millions of small holes in an array like manner. Each hole we fill with nucleotide, one nucleotide at a time.

Microarray II

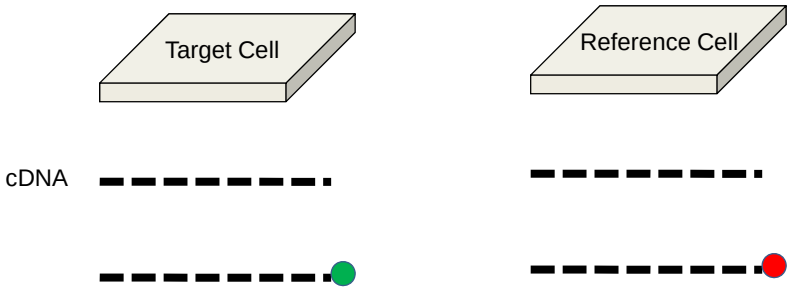




Step 2: Using photolithographic techniques, nucleotides for selected cDNA segments are built up cell by cell creating a collection of binding sites of single stranded DNA sections about 25 nucleotides deep/long on the surface of an NXM array. Each cell becomes sticky for a specific DNA segment.

Step 3: Now we take two DNA samples, one from what we call the Target, the plane we wish to categorize. We then take the segments we collected in step one and tag then with green or red tags, green say for the Target and Red for the Reference.

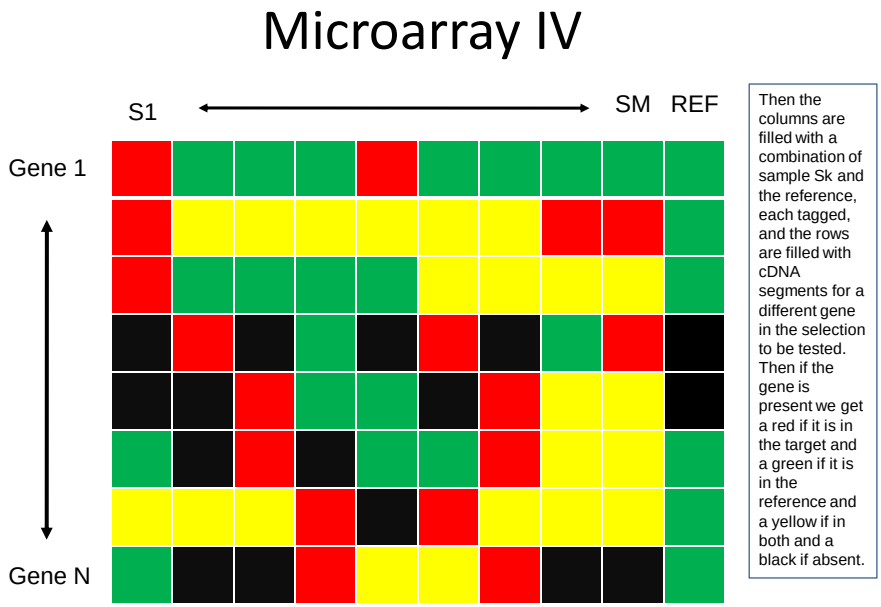
Microarray III



Step 3: For the DNA to be analyzed and a "Reference" target DNA, the mRNA is extracted from each and the cDNA is produced for every gene in the cells to be analyzed, and then it is tagged with a dye which is red for one and green for the other. Typically we tag the target red and Reference green.

Step 4: We then take the samples from the differing plants, one in each column, and look at the array. If the microarray cell has the gene sequence we are seeking to march, and the Target has

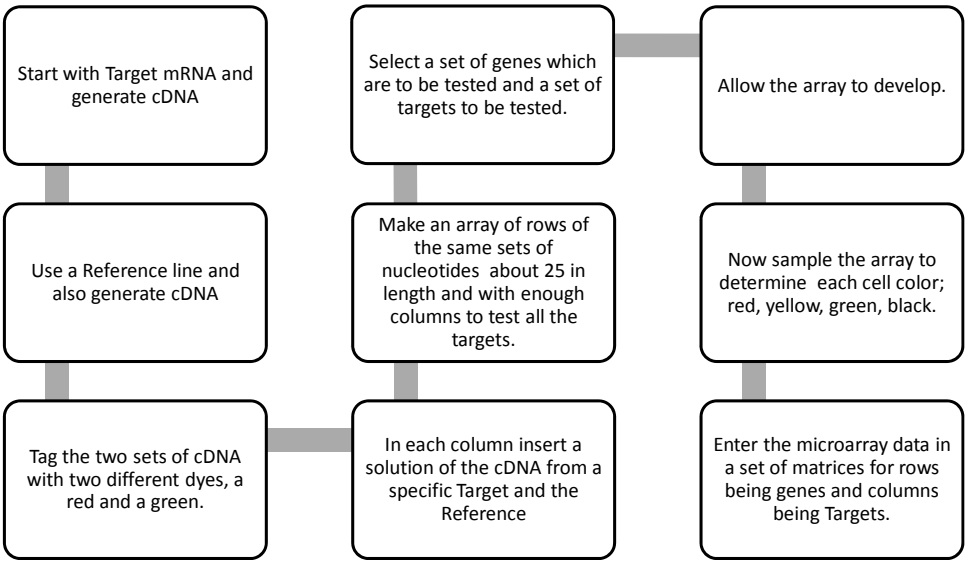
that sequence, it will bond and stick. If the Reference has it, it too will bond. If we just get the Target the cell will be green, if we just get the Reference the cell is red, if we get both the cell turns yellow, and if we have neither the cell is black. The result of a sample scan is shown below.



Now, we even get to try and look at the intensity of the red, green, or yellow. This we can try to see how much is expressed not just whether it is or is not. We will not discuss that here. In the above matrix we can see that many genes are expressed in one or both or none. If we have enough genes than we can argue we have the basis for an exceptionally good means to develop a classification.

In the following Figure we summarize the microarray process.

Microarray Summary



This is a brief summary of microarray technology. We refer the reader to the volume of literature available and referenced to herein.

The Bayesian network approach can be posed as a general problem. Yet it is often delimited by assuming a priori that certain graphs, G, and no possible, and the minimization or maximization process then is performed over the subset of graphs. There is often still quite a great deal of calculation required.

On the negative side the Bayesian approach as is the case for the Boolean approach does not take into account the underlying chemical dynamics that we have been observing. In addition for most of the pathways we are looking at we already know the key pathway structure and it is the reaction kinetics that we are looking for.

Thus Boolean and Bayesian are useful for identifying pathways whereas the full system model works when the pathway is somewhat well understood and we are looking more for the dynamics and details of the reactions. In a sense they are useful at different stages of the process.

4.3 REACTION RATE MODELS

There has been a considerable amount of research on cellular dynamics focused on what is called the reaction rate models. We briefly examine them here and then we will argue that they are inappropriate both specifically and in general for cellular dynamics.

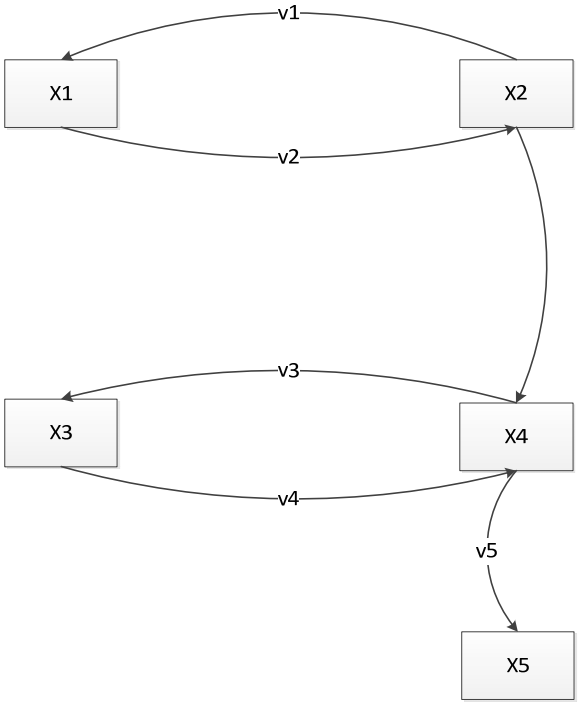
The basic assumptions in the Reaction Rate models are as follows:

1. Concentrations: We assume that we have enough molecules of proteins that we can deal with concentrations rather than molecular interactions. This is the large concentration assumption. It should be noted that this is often if not always violated.

2. Intracellular: We assume that all the action occurs intracellular. We do not really consider intercellular matrices or even receptor-ligan interaction. We believe that the neglect of these factors is a substantial flaw.

3. Rate Reactions: The assumption is that concentrations are looked at temporally, namely rates of change in density, and these rate equations then relate densities of one substance to another via mechanisms of rate reaction theory.

Let us consider a simple model as below where we have five substances, say proteins, and we deal with concentrations and that we now want to develop a model for these concentrations.



Recall that if we have a pathway with say N=5 constituents then we have the following formula for the rates of change of the above described concentrations of proteins where the rates, v, are yet to be determined:

$$\frac{d[x_1(t)]}{dt} = v_1 - v_2$$
$$\frac{d[x_2(t)]}{dt} = v_2 - v_1$$
$$\frac{d[x_3(t)]}{dt} = v_4 - v_3$$
$$\frac{d[x_4(t)]}{dt} = v_3 - v_4$$
$$\frac{d[x_5(t)]}{dt} = v_5$$

Now let us return to the rate analysis. We will use the paper by Segel and Slemrod as the basis. This paper presents an excellent analysis of the quasi steady state assumptions for rate and reaction with enzymatic issues as we will encounter.

Assume we have the following reaction. We assume that E, S, and P are concentrations of proteins.

$$E + S \rightleftarrows C \rightarrow E + P$$

We further assume that the forward rate from E+S to C us k_1 and the reverse is k_{-1} and the forward rate from S to E+P is k_2. Then we have the following:

$$\frac{dE}{dt} = -k_1 ES + k_{-1} C + k_2 C$$
$$\frac{dS}{dt} = -k_1 ES + k_{-1} C$$
$$\frac{dC}{dt} = k_1 ES - k_{-1} C - k_2 C$$
$$\frac{dP}{dt} = k_2 C$$

Note that as we have discussed before the equations are nonlinear in concentrations due to the product terms. We could linearize them by assuming that we are dealing with small changes or we can use the quasi steady state assumptions which reduces the equations yet keeps the nonlinearity. We shall do the later. Also note the mapping of the above specific reactions to the generalized equations we have used above. Also note that when we have two constituents using an enzyme we see the enzyme as a single combine entity and the initial components as two separate ones. Also we have reaction rates thus dependent to the product of the concentrations on each side of the reaction.

Now let us proceed with the enzymatic reaction. We can restrict the analysis to the following conditions:

$$E(0) = E_0$$
$$S(0) = S_0$$
$$C(0) = 0$$
$$P(0) = 0$$

This is a simple defined initial state when we have just the initial entities which we desire to react. But we also can readily show:

$$E(t) + C(t) = E_0$$

Thus we can simplify the result to what we have below by elimination:

$$\frac{dS}{dt} = -k_1(E_0 - C0S + k_{-1}C$$
$$\frac{dC}{dt} = k_1(E_0 - C)S - k_{-1}C$$
$$S(0) = S_0$$
$$C(0) = 0$$

These are still rather complex differential equations and we will now assume that the rates of each are dramatically different. Namely one proceeds quite quickly and the other slowly. That being the case we can thus assume that one is steady state and the other dominates the observable time variation. This is in essence the quasi steady state assumption. This assumption is:

$$\frac{dC}{dt} \approx 0$$

Then we obtain:

$$C = \frac{E_0 S}{K_m + S}$$

where

$$K_m = \frac{k_{-1} + k_2}{k_1}$$

and finally we have the single differential equation for S, which is:

$$\frac{dS}{dt} = \frac{k_2 E_0 S}{K_m + S}$$

This can be applied to the above initial generic form depending upon what the reaction is. Simple reactions do not involve an enzymatic catalyst as we have shown here but in general we will always be dealing with some nonlinearities.

The above deals with four of the five elements we discussed previously: intracellular pathway which was given, rate reactions, concentrations and temporal dynamics.

Thus for any of the pathways we have described above we can readily apply the model we have developed. The issue will be one of simplification. In the previous quasi steady state analysis we eliminate one reaction as not being significant to the "scale" of the reaction or reactions of interest. We assume it has reached a steady state by the time we enter the second reaction. There are two other approaches; first neglect scale and just deal with the complex nonlinearities and compute results or at the other extreme linearize the system.

Let us apply linearization to the above:

$$\frac{dS}{dt} = \frac{k_2 E_0 S}{K_m + S}$$

let

$$S = S_0 + s$$

where

s = *small variation*

then

$$\frac{ds}{dt} = \frac{k_2 E_0 (S_0 + s)}{K_m + S_0 + s} \approx a + bs$$

where

$$a = \frac{k_2 E_0 S_0}{K_m + S_0}$$

$$b = \frac{k_2 E_0}{K_m + S_0}$$

Thus we can look at small variations in such a linear manner.

$g(a) = 0$

is the desired result. Define:

$$A(a) = -\left[\frac{\partial g(a)}{\partial a}\right]^{-1}$$

where we define:

$$\left[\frac{\partial g(a)}{\partial a}\right] = \begin{bmatrix} \frac{\partial g_1}{\partial a_1} & \cdots & \frac{\partial g_1}{\partial a_n} \\ \frac{\partial g_n}{\partial a_1} & \cdots & \frac{\partial g_n}{\partial a_n} \end{bmatrix}$$

and the estimate at sample k+1 is:

$\hat{a}(k+1) = \hat{a}(k) + A(\hat{a}(k))g(\hat{a}(k))$

Note that we use this iterative scheme as one of several means to achieve the result. For each tuple of data we do the following:

$\hat{a}(0) = a^0$, an n x 1 vector guess. Then we use the first data tuple:

$\hat{a}(1) = \hat{a}(0) + A(\hat{a}(0))g((\hat{a}(0))$

where we use the difference:

$a_{k,measured}(0) - \hat{a}(0)$

as the data entry element for each of the elements of a.

The Newton algorithm is but one of many possible algorithms. We know the conditions for Newton convergence. We can also estimate the accuracy of this algorithm as well.

4.4 STATE MODELS

The state dynamic network model for pathways is the heart of what we have been developing herein. It consists of the following:

1. Intracellular Pathways: The description of pathways within the cell and a description which focuses on concentrations of products and exogenously extracellular products that affect the pathways.

2. Intercellular Pathways: This is the modeling of pathways between and amongst the cells. It is intercellular signalling pathways where receptors and their associated ligands are studied. Again this may generally be focused on concentrations of ligands and similar extracellular and intercellular communicators.

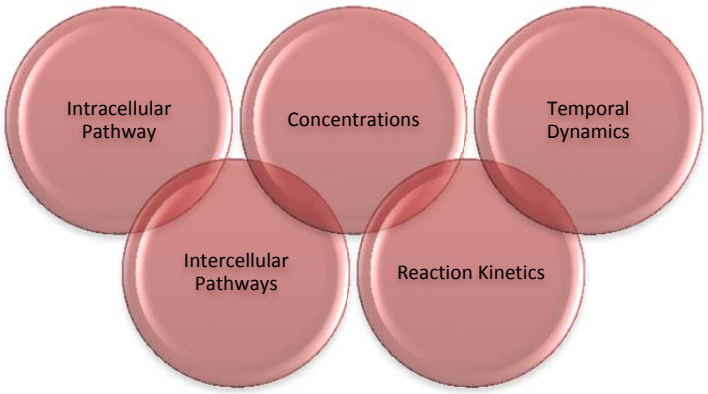
3. Concentrations: The focus is on concentrations of the gene products, but we will assume a looser definition of concentrations, namely a count on specific molecules. The issue here is that it is concentrations which reflect reality and although binding and pathway control may be seen

as promoted or inhibited by other proteins, the reality of reaction kinetics is that it is concentrations which are reflective not just single molecule presence or absence. This element is dramatically different from what we see in Boolean or Bayesian models. Here with the modeling with concentrations we are modeling closer to reality.

4. Reactions: Reaction kinetics describes how one substance can control the conversion of another substance into a third. The reaction kinetics is essential to the system dynamics model. The challenge is to know and understand which of the reaction models apply. For example are there rate limiting factors due to concentration limitations of the reaction. However as we have shown before, we are at so low a concentration in a cell that we may very well be dealing with protein to protein bonding, not looking at concentration based reactions. We have not examined this in detail but it will be considered later.

5. Time Variations: The models are all time varying. Thus understanding reaction kinetics from a temporal perspective is essential. This may take our knowledge to the limit and perhaps beyond at the current time.

We graphically demonstrate these factors below:



In order to best understand the system dynamic model we again briefly return to the reaction rate model and its part in defining the reaction rate equations.

The method of estimating the structural elements of the gene expression can be structured using a standard set of methodologies. In particular we use the two approaches. The approach was applied to estimating the constituent chemical concentrations of the upper atmosphere, namely the inversion problem, using transmitted light as the probe mechanism. In this case we seek to estimate the gene expression matrix using the concentrations of secondary chemicals as expressed in color concentrations. This is in many ways a similar problem.

Let us consider a six gene model, two color modifying genes and four control genes, two each. The model is as follows. First is a general linear model for the gene production:

$$\frac{dx(t)}{dt} = Ax(t) + u(t) + n(t)$$

Then the entries are as follows:

$$A = \begin{bmatrix} a_{11}..a_{12}..a_{13}..0..0..0 \\ 0...a_{22}..0...0...0...0 \\ 0...0...a_{33}..0...0...0 \\ 0...0...0...a_{44}..a_{45}..a_{46} \\ 0...0...0...0...a_{55}..0 \\ 0...0...0...0...0...a_{66} \end{bmatrix}$$

and

$$u(t) = \begin{bmatrix} u_1 \\ ... \\ u_6 \end{bmatrix}$$

And we assume a system noise which is white with the following characteristic:

$$E\left[n(t)\right] = 0$$

and

$$E\left[n(t)n(s)\right] = N_0 I\delta(t-s)$$

Now we can define:

$$A = \begin{bmatrix} A_1...0 \\ 0...A_2 \end{bmatrix}$$

Where we have partitioned the matrix into four submatrices. This shows that each gene and its controller are separate. Now we can determine the concentrations of each protein in steady state as follows, neglecting the Gaussian noise element for the time being:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = -A_1^{-1} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

and

$$\begin{bmatrix} x_4 \\ x_5 \\ x_6 \end{bmatrix} = -A_2^{-1} \begin{bmatrix} u_4 \\ u_5 \\ u_6 \end{bmatrix}$$

We argue that finding either the matrix elements or their inverse relatives is identical. Thus we focus on the inverse elements.

Now we have some measurement model which reflects the concentration of the respective protein. For example we may use a microarray as a way to ascertain concentration. We then could obtain some measurement, say the color of a cell in a microarray, which we call z and then

we can relate the z level to the concentrations x in the cell. The concentrations of the respective proteins as measured by say a color in a microarray are given by the 2 x 2 vector as follows:

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} c_{11}...0...0...0...0...0 \\ 0...0...0...c_{24}...0...0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = Cx$$

If we do not know the A matrices we may append to this model a system identification model. The system identification model is given as follows. Let us begin with a model for the vector a that we seek:

$$\frac{da(t)}{dt} = 0 : where$$

$$a(t) = \begin{bmatrix} a_1 \\ ... \\ a_5 \end{bmatrix}$$

In this case we have assumed a is a 5 x 1 vector but it can be any vector. The measurement system equation is given by:

$$z(t) = g(a,t) + w(t)$$

Where z is an m x 1 vector. In this case however we have for the measurement the following:

$$z(t) = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ x_1 \\ ... \\ x_6 \end{bmatrix} = g(a,t) + w(t)$$

We now expand in a Taylor series the above g function:

$$g(a,t) = g(a_0,t) + C(a_0,t)\left[a(t) - a_0(t)\right] +$$

$$\frac{1}{2}\sum_{i=1}^{N} \gamma_i \left[a - a_0\right]^T F_i \left[a - a_0\right] + ...$$

Where we have:

$$C = \begin{bmatrix} \frac{\partial g_1}{\partial a_1} & \cdots & \frac{\partial g_1}{\partial a_n} \\ .. & .. & .. \\ \frac{\partial g_m}{\partial a_1} & \cdots & \frac{\partial g_m}{\partial a_n} \end{bmatrix}$$

Thus we have for the measurement:

$$z(t) = C(t)a(t) + \left[g(a_0) - C(a_0)a_0(t) \right]$$

We now use standard Kalman theory to determine the mean square estimate;

$$\frac{d\hat{a}(t)}{dt} = P(t)C^T(t)K^{-1}(z - C(t)\hat{a}(t))$$

where

$$\frac{dP(t)}{dt} = -P(t)C^T(t)K^{-1}C(t)P(t) +$$

$$\sum_{i=1}^{N} PF_i P\gamma_i^T K^{-1}(z - g(a_0)$$

where

$$K\gamma(t-s) = E\left[w(t)w^T(s) \right]$$

In discrete time we have the equation:

$$\hat{a}(k+1) = \hat{a}(k) + PCK^{-1}\left[z(k) - \hat{z}(k) \right]$$

This is identical to the equation we would have derived from the Newton steepest-descent method.

The state model above focuses on internal measures. However the model can be expanded conceptually to include extracellular characterizations. As we proceed we do the following:

1. Intracellular: In the next Section we develop an intracellular model. This is a model which includes measures of protein levels, in a manner consistent with what is appropriate, and then develops a classic model for cellular spatio temporal dynamics.

2. Extracellular: In the following Sections we then build upon the first model to use the State Model methodology to develop a model for cell count on an Ensemble basis in a spatio-temporal manner. This second model is the ultimate goal of this approach.

4.5 SUMMARY

We will now summarize the models discussed herein.

We summarize their strengths and weaknesses below.

Model	*Strength*	*Weakness*
Bayesian	This is a data driven model wherein the data generates the connections. It makes optimal use of the microarray technology.	Does not allow intercellular models in a simple manner.
Boolean	Provides a network approach based on a set of on-off switches which models generally the internal pathways and it utilizes the pathways as the core of the model.	It is model driven rather than reality driven.
Reaction Rate	Provides a dynamic intracellular model using a standardized methodology.	Reaction rate theory relies on internal only mechanisms and assumes high concentrations.
State Dynamics	Uses a standard approach for the dynamics of measurable quantities such as number of proteins or number of cells.	The model is quite complex. The model focuses on internal structure but can be readily focused on cellular counts. We complete this task in the next sections.

As we demonstrate above, each of these models has a strength and weakness. We will now develop a fifth model, a model which attempts to combine the best of the above and addresses the complex elements of the spatio-temporal along with the Markovian changes in cellular pathways. The model that we move on to is one which focuses on the cell from an external perspective and includes pathway and protein expression changes as event that modifies the cell as an ensemble.

5 TOTAL CELLULAR DYNAMICS: AN EXAMPLE

The concept of a model of Total Cellular Dynamics is somewhat innovative. It focuses on the movement of the cancer cells throughout the body. In this Section we consider presenting a general model of cancel movement in space and time. We present the model and we will later add to this the reality of genetic mutations that occur as cancers progress.

One of the issues we often deal with in cancer is the issue of stem cells or no stem cells. Simply stated there are certain models which say that cancer is derived from a single stem cell. Then that stem cell collection may proliferate but they represent but a small but controlling faction of the entire collection of malignant cells. When we examine say a prostate cancer then there may be but a small percentage of the malignant cells which are the controlling cells of that cancer. Remove those cells and you kill off the cancer. We do not want at this stage to consider the details of that approach. We suggest that one of the many mutations may very well be that stem cell line. We do not expressly identify it but they are included in our mix. We will consider three possible possibilities:

1. No Stem Cells

2. Stem Cells but Fixed at Initial Location

3. Stem Cells which are mobile.

In Case 1 all malignant cells are clones of each other at least at the start. As the malignant cells continue through mitosis additional mutations are likely so that after a broad set of mitotic divisions we have a somewhat heterogeneous set of malignant cells, some more aggressive than others. As with most such cancer cells they also produce ligand growth factors which stimulate each other and result in the cascade of unlimited growth and duplication.

In Case 2 we assume that there was a single cell which mutated and that this becomes the CSC. The CSC replicates producing one CSC for self-replication and TICs which migrate. We assume that the CSC may from time to time actually double, but not at the mitosis rate of the base. Furthermore we assume the CSC sends out growth factors, GF, to the TICs. The GF flow outward in a wave like manner from the somewhat position stabilized CSCs to the TICs which are mobile and both diffuse and flow throughout the body. The GF must find the TICs which become a distant metastasis.

In Case 3 in contrast to Case 2, we assume mobile CSC and thus the CSCs also flow according to some set of rules.

5.1 TOTAL CELLULAR DYNAMICS MODELS

Now depending on the case we assume we can model the flow of cancer cells according to some simple dynamic distributed models[3]. We now just posit a model almost akin to the Schrodinger approach in Quantum Mechanics. We know that cells diffuse, flow and grow. Thus we create a model which includes each element and we assume some form of linearity. Let us assume we have some parameter p(x,t) which reflects cellular presence at some spatial point x, a multidimensional location, and time t.

Thus we could have:

$$\frac{\partial p(x,t)}{\partial t} = D(x,t)\frac{\partial^2 p(x,t)}{\partial x^2} + E(x,t)\frac{\partial p(x,t)}{\partial x} + F(x,t)p(x,t)$$

This provides diffusion, flow, and rate elements. The rate term, the F term, is a rate of change in time at a certain location and time specific. It is the duplication rate at that specific location due to the normal mitotic change. The last term may be both pathway and environment driven.

We can argue from many other physical phenomena that this model is the basis of any spatio-temporal changes.

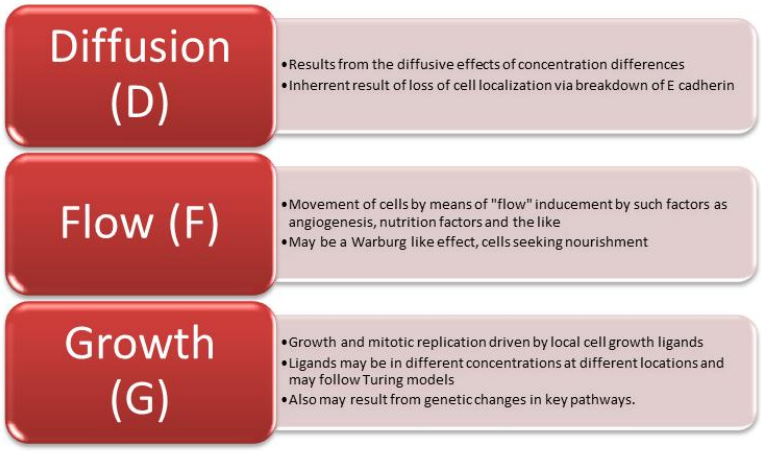
1. Diffusion is a typical process that occurs as we increase density in one region there is a natural tendency to create a balance of density. Specifically increase density tends to force movement in a diffusive manner from an area high density to low density. The diffusion constant D represents the characteristics of the cell of that specific type in a region of a specific type. It is important to note that D is both cell and environment related.

2. Flow is the movement of cells due to some flow type process. Namely cell can be "pushed" along by many factors. They move in the blood stream, the move in the lungs and kidneys.

3. Growth is the final term. This is the growth factor F which is also cell and region dependent. Certain cells may find a rich environment to grow, such as the liver or bone.

Now this description has certain physical realities.

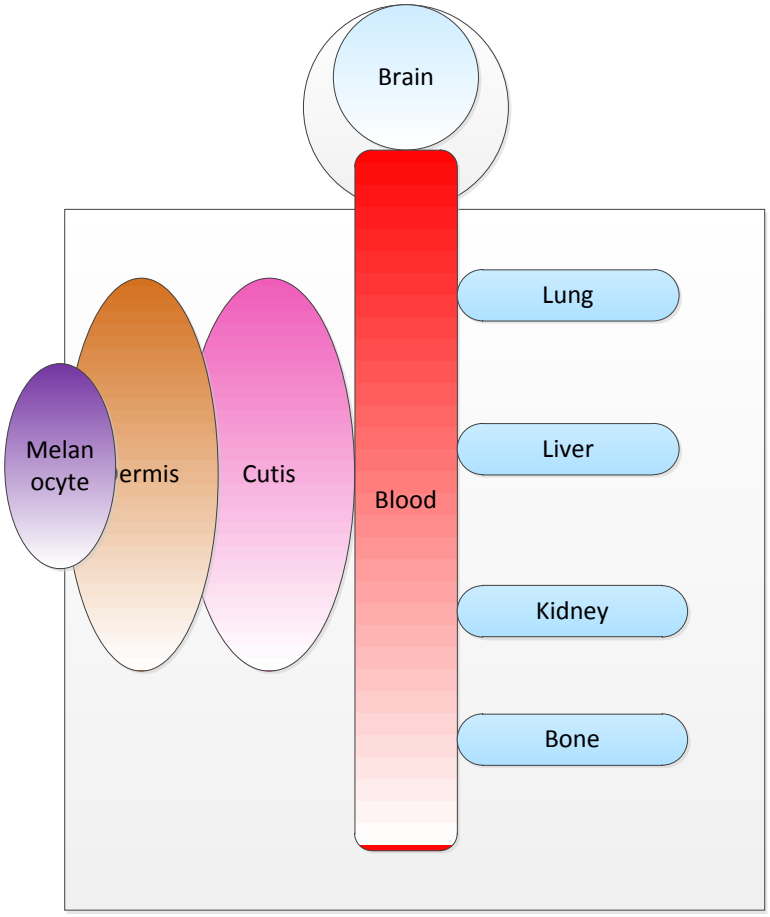
[3] See Andersen p 277 of Bellomo et al for a variant on what we are proposing here. The Andersen model is somewhat similar but lacks the detail we present herein. Also there is in the same volume a paper by Pepper and Lolas focusing on the dynamics of the lymphatic cancer system, p 255.



Here above we describe the three factors in terms of their effects and their causes. The three elements of the equation; diffusion, flow, and growth, are the three ways in which cancer cells move. We can summarize these as below:

Factor	***Diffusion***	***Flow***	***Growth***
Physical Effect	Cancer cells begin to diffuse due to concentration effects.	Cancer cells are "forced" to move by a flow mechanism driven them in a direction along flow lines.	Cancer cells begin to go through mitosis and cell growth.
Genetic Driver	Movement is due to the loss of location restrictors such as E cadherin found in melanocytes and restricting their movement.	Flow lines may be developed by means of metabolic needs of the cell in search of the nutrients required for growth. This may be a combination of angiogenesis as well as a Warburg like effect.	Growth factor ligands attach to the surface of the cell. Flow of such ligands and their production may be influenced by a Turing flow effect thus accounting for complexity of location of growth.
Impact	Slow migration in local areas.	Cells have lost functionality and move to maximize their nutrition input to facilitate growth.	Cancer cells may find optimal areas for proliferation based upon factor related to ligand density.

Now consider the following graphic as a human body. Here we show an example of a melanocyte that may become malignant and the other general regions where it may move and grow in.

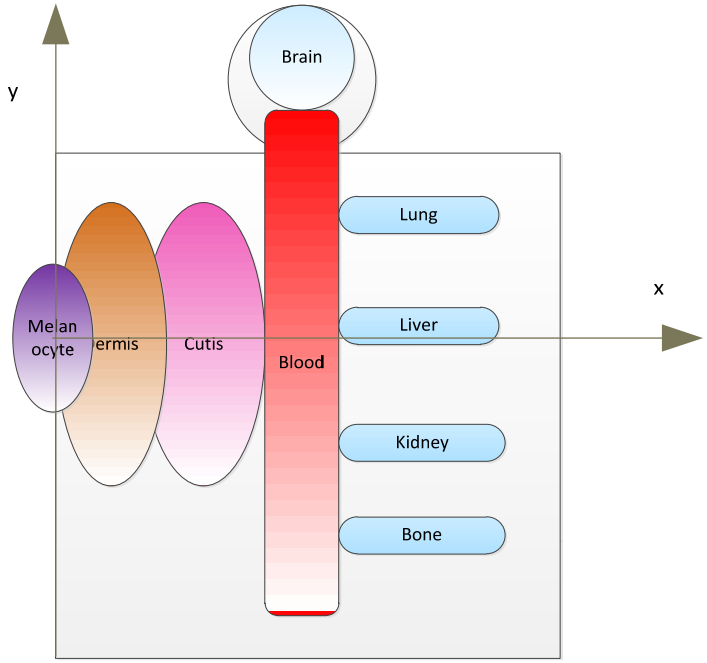


We have a D, E, F, for each gross portion of the above simplified body. We also have a model as specifically below in the Table:

Organ	*D* *Diffusion*	*E* *Flow*	*F* *Production*
Epidermis	0.5	0.01	0.7
Dermis	0.4	0.02	0.5
Cutis	0.3	0.05	0.2
Blood	5.0	0.5	0.01
Brain	0.1	0.01	0.2
Liver	2.0	0.2	0.3
Lung	3.0	0.3	0.4
Kidney	1.5	0.4	0.5
Bone	2.5	0.5	1.0

The above numbers are purely speculative. But if we can ascertain them then we get a solution of p(x,t) in time. Note that here we have a two dimensional space. Thus we have the above constants applying only to this artifactually spatial model. Distance is measured in terms of

movement across the interfaces. For simplicity we assume that all other space is impenetrable by any means. This we have production, flow and diffusion in each area.



Note that in the above we have laid out the x and y coordinates such that we have blood flow in the center, namely the metastasis flows via blood, and then enters organs as shown. The "location" of the organs is distances. Note also the origin of the malignancy is at (0,0).

Now we can relate the constants to the pathway distortions which are part of the malignancy as well.

The question is; how do we determine these constants so that we may verify the model? Let us assume we can do so via examination of prior malignancy, not an obvious task but one we shall demonstrate. One must be cautious also to include in the determination pathway factors for each malignancy and its state and stage. Thus the three constants will be highly dependent upon the specific genetic makeup of the initial malignancy.

5.2 TURING TESSELLATION

In 1952 Alan Turing, in the last year and a half of his life, was focusing on biological models and moving away from his seminal efforts in encryption and computers. It was Turing who in the Second World War managed to break many of the German codes on Ultra and who also created the paradigm for computers which we use today. In his last efforts before his untimely suicide Turing looked at the problem of patterning in plants and animals. This was done at the same time Watson and Crick were working on the gene and DNA. Turing had no detailed model to work

with, he had no gene, and he had just a gestalt, if you will, to model this issue. Today we have the details of the model to fill in the gaps in the Turing model.

The Turing model was quite simple. It stated that there was some chemical, and a concentration of that chemical, call it C, which was the determinant of a color. Consider the case of a zebra and its hair. If C were above a certain level the hair was black and if below that level the hair was white. As Turing states in the abstract of the paper:

"It is suggested that a system of chemical substances, called morphogens, reacting together and diffusing through a tissue, is adequate to account for the main phenomena of morphogenesis. Such a system, although it may originally be quite homogeneous, may later develop a pattern or structure due to an instability of the homogeneous equilibrium, which is triggered off by random disturbances. Such reaction-diffusion systems are considered in some detail in the case of an isolated ring of cells, a mathematically convenient, though biologically unusual system.

The investigation is chiefly concerned with the onset of instability. It is found that there are six essentially different forms which this may take. In the most interesting form stationary waves appear on the ring. It is suggested that this might account, for instance, for the tentacle patterns on Hydra and for whorled leaves. A system of reactions and diffusion on a sphere is also considered. Such a system appears to account for gastrulation. Another reaction system in two dimensions gives rise to patterns reminiscent of dappling. It is also suggested that stationary waves in two dimensions could account for the phenomena of phyllotaxis.

The purpose of this paper is to discuss a possible mechanism by which the genes of a zygote may determine the anatomical structure of the resulting organism. The theory does not make any new hypotheses; it merely suggests that certain well-known physical laws are sufficient to account for many of the facts. The full understanding of the paper requires a good knowledge of mathematics, some biology, and some elementary chemistry. Since readers cannot be expected to be experts in all of these subjects, a number of elementary facts are explained, which can be found in text-books, but whose omission would make the paper difficult reading."

Now, Turing reasoned that this chemical, what he called the morphogen, could be generated and could flow out to other cells and in from other cells. Thus focusing on one cell he could create a model across space and time to lay out the concentration of this chemical. He simply postulated that the rate of change of this chemical in time was equal to two factors; first the use of the chemical in the cell, such as a catalyst in a reaction or even part of the reaction, and second, the flow in or out of the cell. The following equation is a statement of Turing's observation.

$$\frac{\partial C_1(x,t)}{\partial t} = F_1(C_1, C_2, x, t) + \lambda_1 \nabla^2 C_1(x,t)$$

This is the nonlinear diffusion equation. It allows one to solve for a concentration, C, as a function of time and space. It requires two things. First is the diffusion coefficient to and from cells and second the functional relationship which shows how the chemical is used within a cell.

5.3 DETERMINING THE COEFFICIENTS

The question now is; how does one determine and link the coefficients in the models? For example if we believe that diffusion D depends on E cadherin concentration, namely as E cadherin decreases then D increases we may postulate a model as follows:

$$D(x,t) = \kappa(\beta - n_{ECadherin}(x,t))$$

In the above, the constants are to be determined by some experimental analysis. We know that the more E cadherin the stickier is the cell and the less diffusion that occurs. Thus the above is at the least a first order approximation.

In a similar manner we can relate F to the concentration of PTEN and p53. We do so as follows:

$$F(x,t) = \kappa_{PTEN}(\beta_{PTEN} - n_{PTEN}(x,t)) + \kappa_{p53}(\beta_{p53} - n_{p53}(x,t))$$

This is merely a suppositional model. However we generally do know the following:

1. The genes which are expressed for adhesion and replication are known.

2. We know the pathways for these genes

3. We know the intracellular models controlling these genes.

4. We know that functionally an excess or paucity of a gene has a certain effect.

5. We know that in general in small amounts the world is linear.

6. We know that we can use regression techniques based upon collected data to determine coefficients in a general sense.

Thus we have a fundamental basis to express the following set of relationships between the coefficients and the concentrations of certain proteins:

$$D(x,t) = \sum_{i=1}^{N} \kappa_i [\beta_i - n_i(x,t)]$$

$$F(x,t) = \sum_{i=1}^{N} \lambda_i [\alpha_i - n_i(x,t)]$$

$$G(x,t) = \sum_{i=1}^{N} \mu_i [\gamma_i - n_i(x,t)]$$

In the above we have described the D, F, and G as linear functions of concentrations of proteins, or other measurable markers as well, and we state that there are 6N constants that must be determined.

Now we have related intracellular concentrations, which themselves may be temporally and spatially dependent, to the total parameter values for the flow of cells throughout the body. We may also want to relate these to organ specific parameters as well.

Thus what we have achieved is as follows:

1. Model relating intracellular and whole body.

2. Methodology to determine the constants.

3. Methodology to go from patient data to prognostic data.

4. Methodologies to establish possible treatment methodologies. Namely what gene controls will result in what whole body reactions.

Now we proceed to provide some detailed substance to this p(x,t) variable. We will develop a model of cell density by cell type, type being defined by a specific genetic expression mode. We will also go and consider the specific genetic changes based upon the internal networks of the pathways in a cell, thus incorporating the genetic dynamics of the cell as it changes.

6 THE COMBINED MODEL

We now want to examine the principles we have used in the model to this point. There are three factors we can use to better effect results or determine what the results could be. The three areas are:

1. Stochastic Models: These are models where we have effects that perturb then which are just random factors. Random in our sense of the word means factors which are by their very nature uncertain. This does not mean unknown, but factors like Brownian motion effects. The Brownian motion factors may then make the cell propagation equations random fields and not purely deterministic.

2. Stability: In control theory we know that there are certain system models which result in unstable systems. They have stability problems, they go out of control. The issue then is cancer a stability problem and if so what are the factors which we can use to stabilize them? How do we characterize stability in a random or stochastic random field?

3. Controllability and Observability: Again in the area of control theory we know that we can examine a system and assess two factors. First is a system Observable, namely if we are told where it is at a certain time and spatial point can we, knowing the system dynamics, ascertain where it will be at some other moment and location? In the case of Controllable, we ask; does there exist some control function which will drive the system to some desired state. These two factors are critical for using a system which we demand to be predictive.

Let us now examine some of these issues in detail.

6.1 STOCHASTIC EFFECTS

Stochastic Models are at the heart of understanding how the systems may function under the following circumstances:

1. When truly random factors influence a pathway. There are true random effects. They are akin to white noise or random noise in a radio signal. They are just artifacts of things that have nothing to do with our known system. They may be of a small nature, thus Gaussian in nature, with no big events, or event like, big factors, occurring from time to time, and thus may be Poisson like. Can we measure them, possibly, can we develop a physical model for them, most likely, but for the most part they take no part in the underlying pathway and they are independent totally, and in turn in a statistical manner.

2. Unknown or Known Effects: Let us consider here the miRNA model as an example. We know they exist, we know they disturb pathways, but we do not have a good model for them. Thus we consider them to be random, and we attempt to provide a model based upon measured reality to incorporate their effects. There are positives and negatives with these approaches.

An alternative approach is to include the unknown/known effects as states to be identified, using classic identification techniques. We have employed that approach from time to time but it may suffer from certain estimation instabilities. We shall examine that later.

The stochastic effects that we may consider are the following:

1. Variations in constants: The various constants that we have assumed in the model may vary randomly. Not only may the linearization result in certain perturbations but there may be actual physical phenomenon that causes changes, ranging from pH to density of cells.

2. Variation in Exogenous Media: We can see that there are random variations in the extracellular matrix and the blood environment in the flow of cells. Also there may be local cellular variations from a mean level depending upon such factors as adipose tissue and the like.

The list can go on. But the most significant stochastic issue is the change in gene expression, loss of function and the like. We shall consider this later.

Thus far the following model has stochastic effects:

$$\frac{\partial p(x,t)}{\partial t} = D(x,t)\frac{\partial^2 p(x,t)}{\partial x^2} + E(x,t)\frac{\partial p(x,t)}{\partial x} + F(x,t)p(x,t) + n(t)$$

Each of the coefficients may be random, and there may be a random driving function n(t). We must be able to deal with this phenomenon as we proceed.

6.2 STABILITY

With highly complex systems, even just linear ones, and more so with nonlinear ones, we often have significant issues regarding their stability. The stability issue may be model related or more interestingly it may reflect the nature of that specific pathway.

The systems we see developed may perforce of their complexity have substantial stability problems. We must be cautious to be assured that the instability if present reflects nature and not the model.

Let us first define what we mean by stability. It means that the system state continues to grow or expand over time. Cancer is a classic example. Cells grow in an uncontrolled manner. Thus we may ask again about the model:

$$\frac{\partial p(x,t)}{\partial t} = D(x,t)\frac{\partial^2 p(x,t)}{\partial x^2} + E(x,t)\frac{\partial p(x,t)}{\partial x} + F(x,t)p(x,t)$$

What are the bounds on the D, E, and F that result in stable homeostasis and what will result in unstable growth. There is a wealth of literature here that can help us and it all depends on how well we can arrive at a model as shown above.

6.3 CONTROLLABLE AND OBSERVABLE

In systems we often are concerned about two key concepts; controllability and observability. Let us first define them and then discuss their importance.

For both cases we assume we have a system of the following type:

$$\frac{dn(t)}{dt} = A(t)n(t) + B(t)u(t)$$
$$z(t) = C(t)n(t)$$

Namely we have some system which has a large vector of proteins in a cell, and that we know the dynamics of these proteins and that we have some control vector u which can control the number of proteins in a cell. Furthermore we assume we can observe the number of proteins via some system which produces a measurement z.

We now introduce the two concepts. We rely upon McGarty (1974, pp 33-41) for the theoretical background

6.3.1 Controllability

Now a system is said to be controllable is we can find some u(t) which allows us to drive the system to some state x(T) and time T. We may not find the specific u immediately but we are assured that one exists.

This is a critical concept since it allows us to say develop a protocol to treat some pathway disorder by selecting some control u, say a selection of a kinase inhibitor. On the other hand if the system is not controllable that negative result may force us to expand the system or try and find an alternative set of controls by changing say the C matrix.

6.3.2 Observability

A system is observable if from a set of measurements, z(t), where tε(r,s), we can determine x(T) for some T. Namely we assume we know the dynamics of the system and that we have measurements over some interval and from these measurements we can then determine x(T) and in turn any other value of x since we know the governing equations. In our case at hand, we assume we have say some microarray or FISH measurements, or others, and that from these we can then determine all of the protein counts at some T and thus at any t. Namely the system will be fully determined.

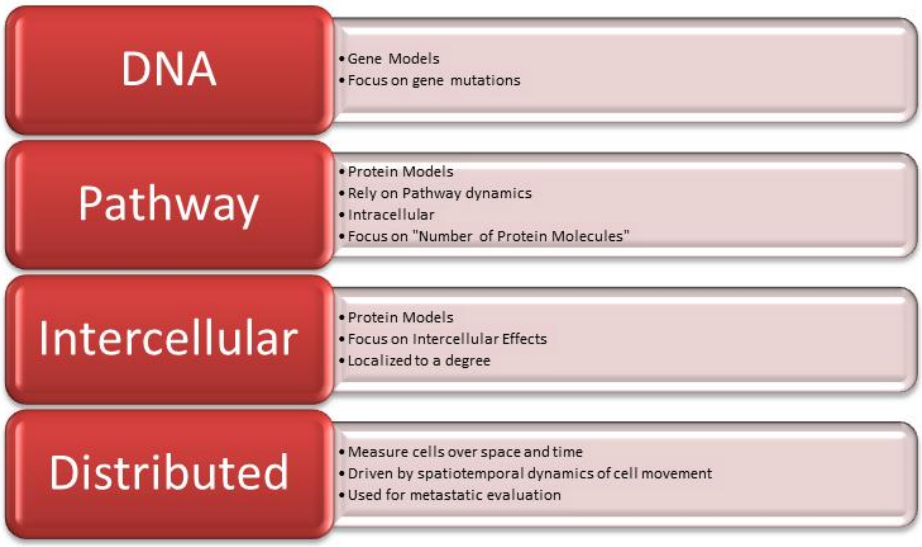
Thus with these two concepts we have a model for a system which can be used to both identify all states and drive it to a desired state.

7 DEVELOPMENT OF ENSEMBLES

We can now summarize this models we have considered. First we should emphasize that for the most part those working in the field have developed pathway models which exhibit a non-temporal mode, it is some steady state model, and the model assumes a protein to protein connection, as if there were a single protein molecule produced and that the interacting proteins were there or not. Part of the simplicity of the models is determined by the limits of what can be measured. We have herein attempted not to limit the results by what can be accomplished currently but has extended the model to levels which assist in a fuller representation of reality. However even here we may very be falling short.

For we have deliberately neglected such things as miRNA, methylation, and the stem cell paradigm just to name a few.

We combine all four methods in a graphic below. We summarize the key differences and differentiators. Currently most of the analytical models focus on pathways. This can generally be supported by means of microarray technology and even rough estimates of relative concentrations may be inferred by such an approach.



The risks we see even in the above models are the absence of exogenous epigenetic factors and the inclusion of a stem cell model. The latter issue is one of major concern. For example if we have true cancer stem cells, CSC, then we have a proliferation of differing cell types. The use of microarrays is for the most part and averaging methodology, not a cell by cell methodology. If we collect cells from say a melanoma tumor. how much of that is a CSC and how much a TIC. And frankly should we identify CSCs only and perform our analysis on those cells alone.

7.1 BASIC PRINCIPLES

We have previously introduced a cancer cell propagation model in earlier discussions which others have also considered. However our model is for a single cell type which proliferates diffuses and flows. The rates of each are dependent on where the cell is. Namely the rates of proliferation, diffusion, and flow depend on the external environmental factors. Thus we often find that proliferation may be enhanced when a cancer cell has metastasized to the liver or brain. The cancer cell may just flow and not proliferate in the blood stream. And the cancer cell may just diffuse at it moves through the vascular wall. The environment is a key determinant.

We now add other elements, namely the probability that a cell can mutate and that as it mutates the factors related to the propagation model may also change. We know that cancer cells do not just have one mutation, but a process of such mutations. The cell may start with a specific change, such as loss of E cadherin for a melanocyte which then allows the cell to move from the basal layer. This may result in a melanoma in situ.

Then we get a mutation in BRAF which allows for proliferation of the movable melanocytes and then loss of p53 for example. Thus there may be a progression of genetic or epigenetic changes in the cell. We now develop a Markov model for this progression, and then we identify collections of cells with the same mutations and apply the same proliferation, flow, and diffusion to each. We calculate a similar diffusion equation now for the average number of malignant cells by region and by type.

7.2 ADDED COMPLEXITY

The previous model described a single mutation. This expands the model by addressing multiple mutations in a Markov manner. That is we demonstrate:

1. The standard diffusion-flow-proliferation model applies on a per-region and per cell type basis. This means that the constants we have developed previously will depend on the specific cell type as well, namely how many mutations have occurred.

2. That we know there are multiple mutations in cancer cells. Some may have a few and are indolent and others may have many and be aggressive. We develop a Markov model for such cell progression.

3. We combine the three element spatio-temporal model with the Markov cell mutation model and this allows us to determine the average number of cells of a specific type in any part of the body at any point in time.

4. We then discuss how one may use this model for prognostic and therapeutic purposes.

The main observation in this brief section is that the average number of malignant cells of a specific mutation state can be determined by the following equation:

$$\frac{\partial \overline{n(x,t)}}{\partial t} = \tilde{L}\overline{n(x,t)} + \Lambda \overline{n(x,t)}$$

In this equation the n(x,t) is an NX1 vector of average numbers in spatio-temporal dependent values of each of N possible genetic expressions or mutations and the L value is the spatio-temporal dependent operator matrix and Λ is a matrix describing the Markov transition probabilities between genetic states or mutations.

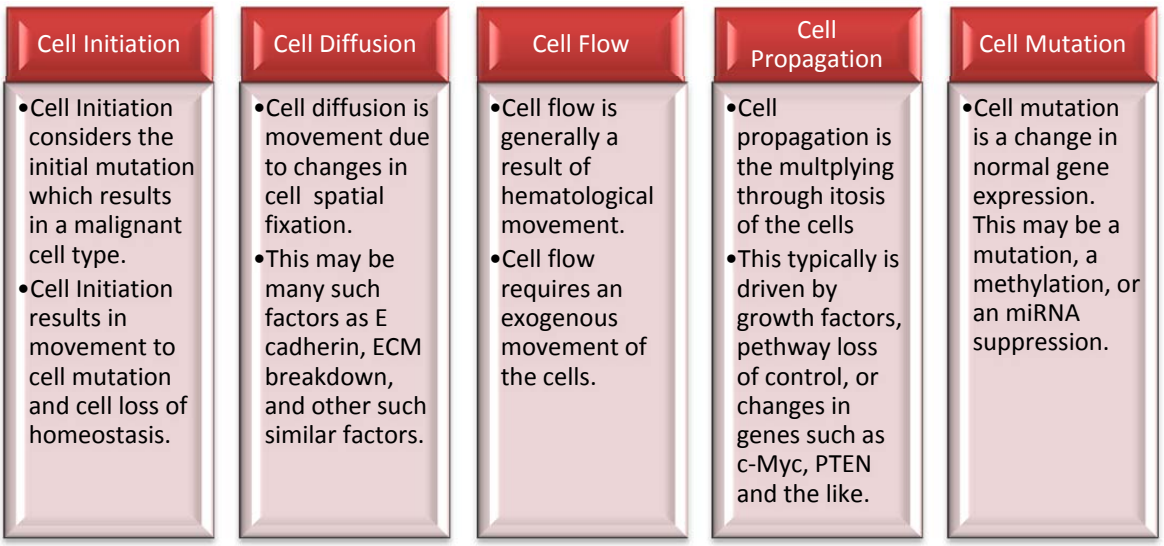
For example, n(x,t) can be thought of as follows:

$$\begin{bmatrix} n_1(x,t) \\ n_2(x,t) \\ n_3(x,t) \\ n_4(x,t) \\ \dots \\ n_N(x,t) \end{bmatrix} = \begin{bmatrix} E(\text{No. cells with normal genetic network}) \\ E(\text{No. cells with BRAF V600 genetic network}) \\ E(\text{No. cells with above and MEK genetic network}) \\ E(\text{No. cells with above and, PI3K genetic network}) \\ \dots \\ \text{E(No. cells with above plus p53 genetic network)} \end{bmatrix}$$

Note in the above we stipulate a specific genetic structure common for the collection or ensemble of cells being counted in that specific spatio-temporal average.

It should be clear that we can measure all of the constants involved and thus determine the result. We have described previously how one could accomplish that determination. As a counter-distinction we can measure the n values and mutation states and determine the constants.

The expanded model considers the issue diagrammed below:



The next issue is the ability to determine what the factors are in the specific model, namely the values of the constants, and secondly the validation of the model itself.

7.3 KEY ISSUES

Thus there are two dimensions of key issues here:

1. Model Identification and Validation: In previous work we referred to this as the Observability problem. Namely if we have a model and we can identify the required parameters, then can this model be used to determine the end state which will be attained. This is the prognostic problem.

2. Model Utilization: As with the previous cases, if we have this model, and we have identified the constants, can we determine actions which may be taken to control the end state of the system? This is the Controllability problem. It states that perhaps having such a model we can determine methods and means to drive the system, in this case the average number of malignant cells of genotype say G, to a new end state, one where we have reduced the number of bad cells to a de minimis level. This is the therapeutic problem.

There also is a third element:

3. Identification: In both of the two previous issues we assumed that there existed a method by which we could determine the constants of diffusion et al and furthermore that we could ascertain the list of possible mutations, and also their Markov transition probabilities. This may be accomplished in two ways. First, we can accomplish this by in vitro studies. Second, we can achieve this by using the model itself in a classic system identification model with in vivo analyses.

Thus the analysis contained herein is an initiation of what appears to be an innovative way to look at cancer. There have been many studies in more specific and segmented areas but there has not to my knowledge been a study that has examined cancer in such a broad and overarching manner. In essence we have included all of the variables that one may ask for.

7.4 UNDERSTANDING METASTASIS

This model is one which attempts to understand metastasis from two dimensions. First, we have examined the movement of malignant cells around the body. The movement and proliferation is driven by the cell dynamics such as the mitotic cycle and its control and the loss of cell specificity and spatial stability.

Metastasis is a somewhat unique characteristic of cancer. Viral diseases, such as HPV and similar viruses, which cause warts, are highly localized. They cause proliferation but localization is maintained.

Finally, recall that with metastasis we have the following typical sites[4]:

Cancer type	Main sites of metastasis*
Bladder	Bone, liver, lung
Breast	Bone, brain, liver, lung
Colorectal	Liver, lung, peritoneum
Kidney	Adrenal gland, bone, brain, liver, lung
Lung	Adrenal gland, bone, brain, liver, other lung
Melanoma	Bone, brain, liver, lung, skin/muscle
Ovary	Liver, lung, peritoneum
Pancreas	Liver, lung, peritoneum
Prostate	Adrenal gland, bone, liver, lung
Stomach	Liver, lung, peritoneum
Thyroid	Bone, liver, lung
Uterus	Bone, liver, lung, peritoneum, vagina

Thus for each of the above sites we have a local cellular and extracellular environment which supports the metastatic behavior that we often see in such cells.

Finally we discuss some of the issue of how do we ascertain the constants in each of the models and this includes the Markov transition probabilities. We examine several approaches, invasive and non-invasive ones. We believe that molecular functional imaging, MFI, provides an attractive approach to ascertaining these constants.

[4] http://www.cancer.gov/cancertopics/factsheet/Sites-Types/metastatic?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+ncifactsheets+%28NCI+Fact+Sheets%29

8 ENSEMBLE MODELS

We have previously developed a simple model for the change in the number of cells of a specific type at a specific place and time as follows.

8.1 THE MODEL

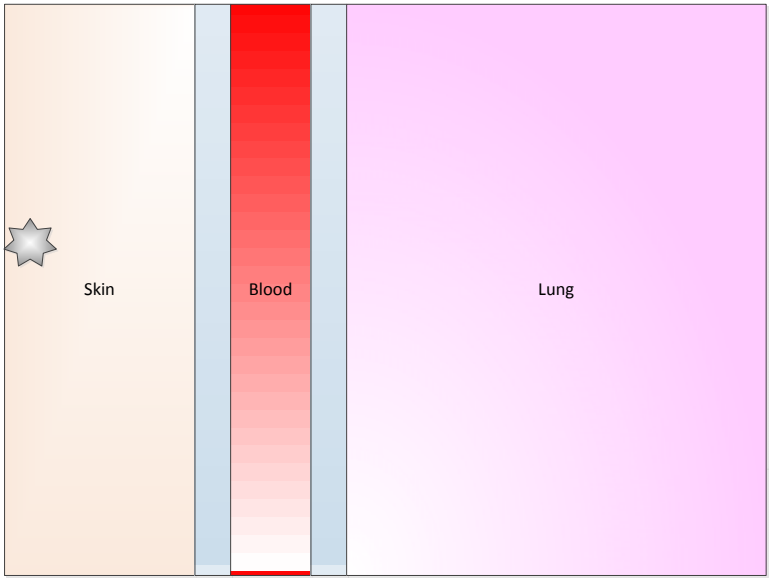
We have demonstrated earlier that for a specific type of single mutated cell that the number of such malignant cells at a specific time t and place x are determined by n(x,t) and that this can be described by the following equation.

$$\frac{\partial n(x,t)}{\partial t} = a\frac{\partial^2 n(x,t)}{\partial x^2} + b\frac{\partial n(x,t)}{\partial x} + cn(x,t)$$

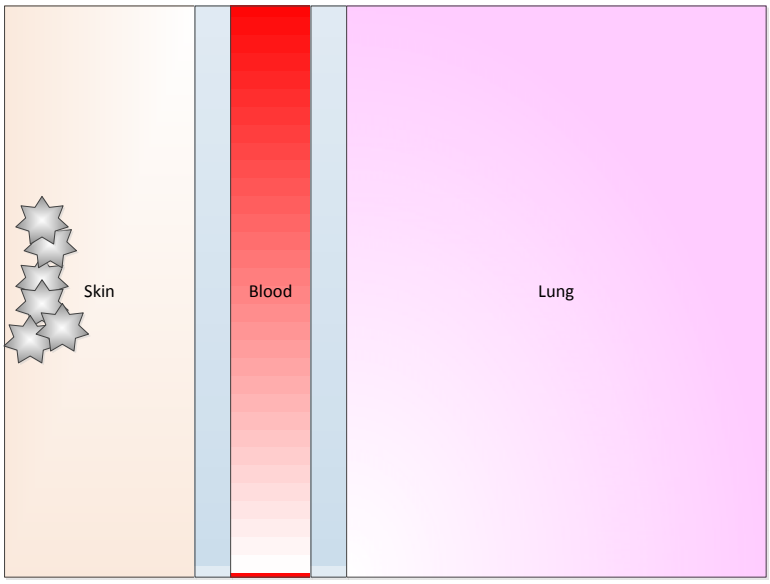
This depicts; diffusion, flow, and growth. The coefficients are cell type dependent and may also be spatially and temporally dependent also. We need not worry about that at this time.

To better understand we depict the progression of melanoma below with a simple graphical example. We start with a simple benign cell, assume a single malignant change and then follow the proliferation and movement of the cell. This graphic makes many simplified assumptions which we shall release shortly..

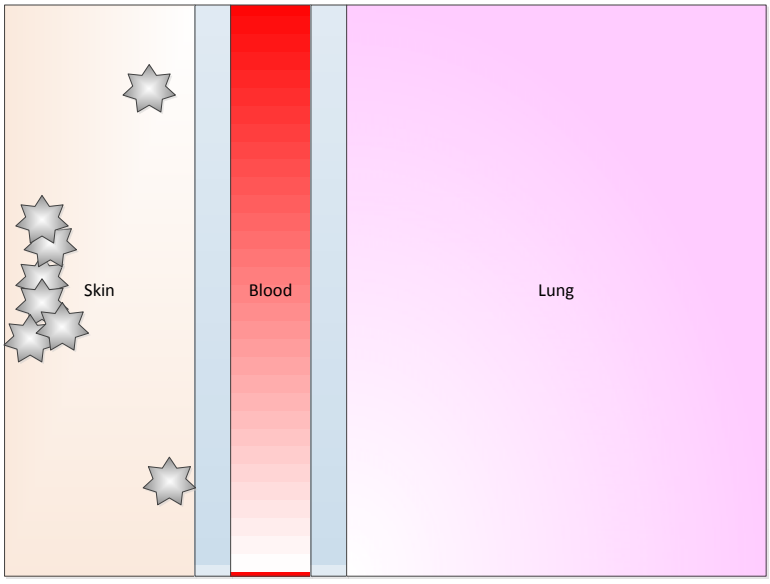
Step 1: Benign State, here we have five segments; skin, two tissue-blood barriers, blood, and lung. We begin by showing a single melanocyte. We assume the melanocyte is affixed to the basal layer with E cadherin functioning properly.



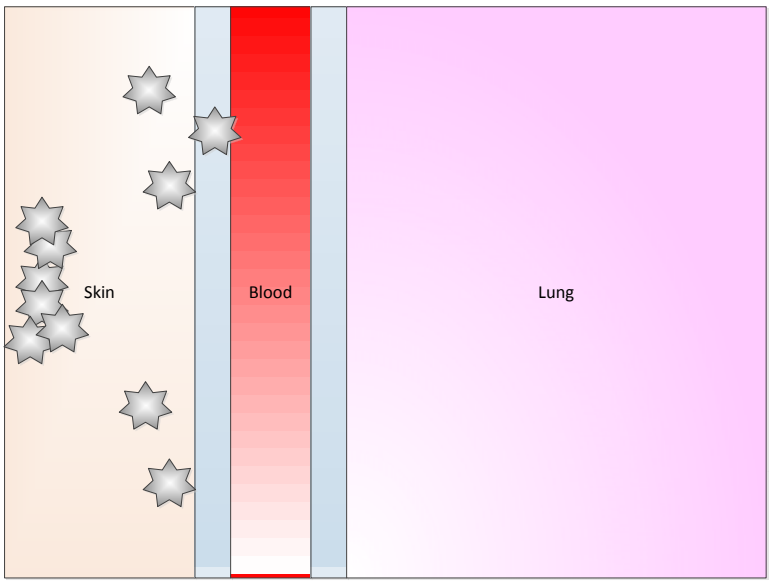
Step 2: We have the beginning of a cancer due to some mutation of the basal or luminal cells. The cancer initially proliferates and then it diffuses. In the figure below we show that it is still localized.



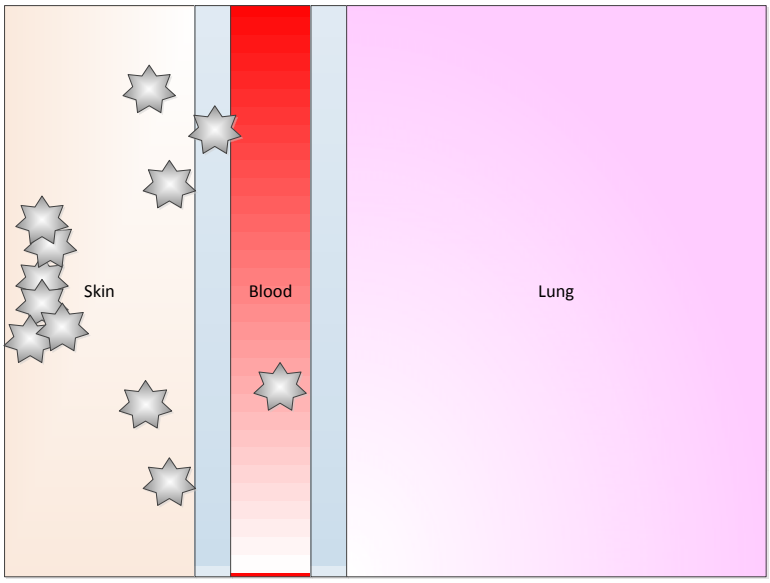
Step 3: Diffusions begins and starts to send the cancer cells towards the blood barrier.



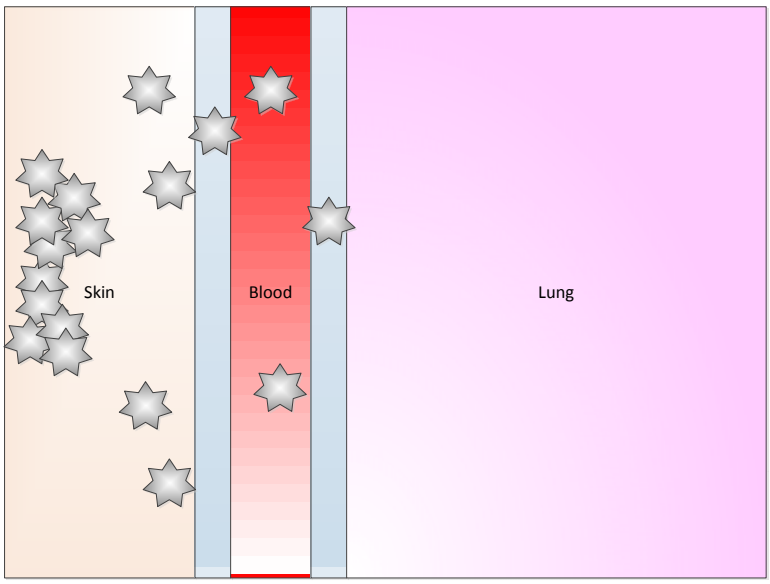
Step 4: The blood barrier is crossed, and we assume by diffusion. Across this barrier there is no proliferation or flow, just diffusion.



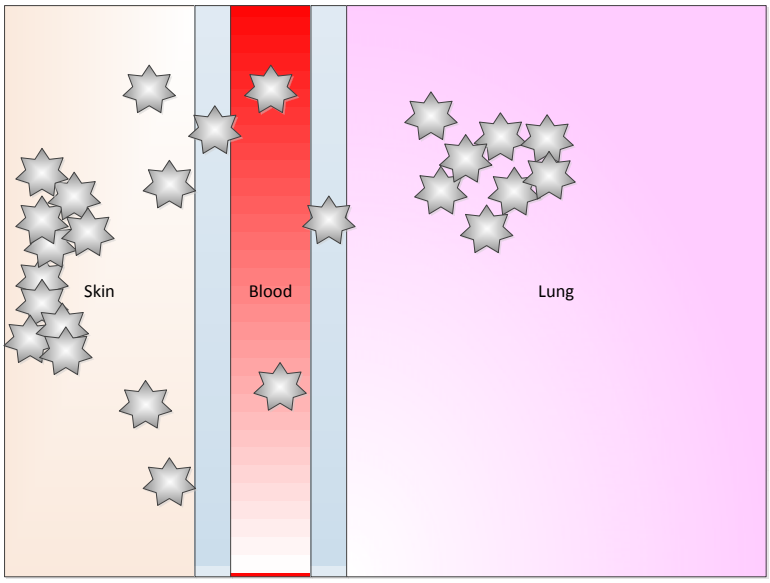
Step 5: The blood barrier is crossed and the cell is now in the blood stream. Here we have flow but no diffusion and no proliferation.



Step 6: The blood barrier is crossed again as discussed above.



Step 7: Metastasis is complete by having the new malignant cells in the lung and proliferation and diffusion predominate.



The above steps are common is almost all cancers. The assumptions here are:

1. The same malignant cell moves across the body.

2. Each separate area, in this case five, has constant diffusion, flow and proliferation constants.

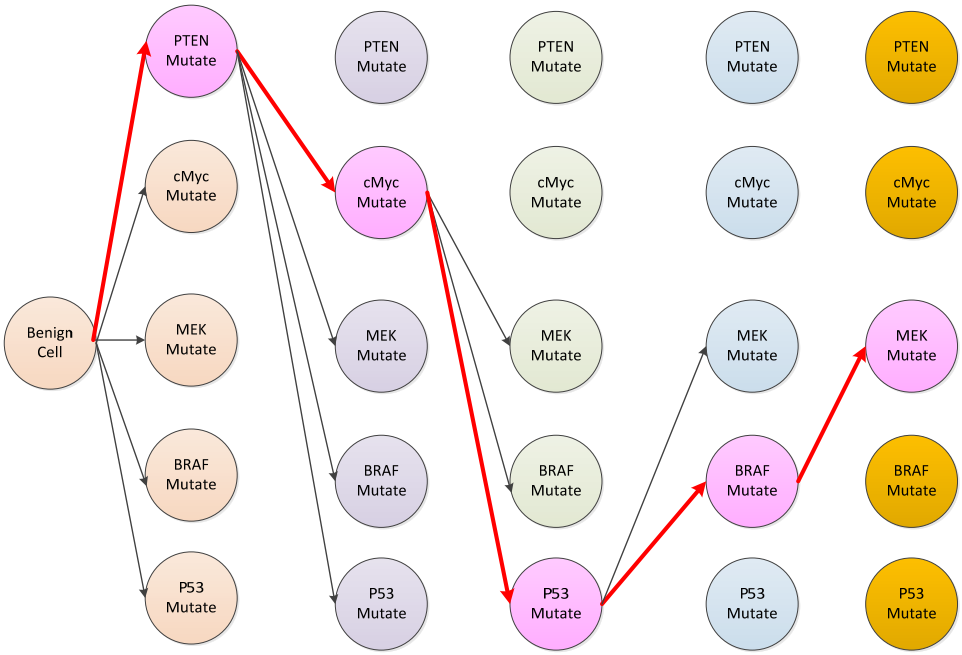
3. That we can then measure the number of cells from this deterministic model.

In the case where they are uniform constants we can solve the equation. In the case where they are uniform constants across uniform spatial domains then we can also solve the equations evoking boundary conditions.

We now want to expand this model to include multiple malignant cell types. Also we want to include their stochastic dynamics as well.

8.2 MARKOV MODELS

Consider a cell with five possible mutations. We show the genes below. The call may begin with one mutation and then move to a second and so forth. Each path is assumed to be possible and the results of each path are different.



Now we can consider a model for the above simple example. We have 5 possible mutations and they may occur in any order. We assume they occur one at a time. We can identify any number of cells as:

$$n_{1,k}(x,t)$$

As the number of cells after one mutation at location x and at time t, of mutation k.

8.2.1 The Markov Assumptions

Now we have the following observations:

1. At mutation 1 we have 5 possible cell mutants. Furthermore each may be considered a cancer cell and the growth, diffusion and flow are as described above. Some of the mutations may be indolent and some aggressive.

2. At mutation 2 we have 5*4 possible cells. The question is that some are say PTEN then cMyc or cMyc then PTEN. Are they the same, and this means the difference between perturbation and combination? Are they distinct by have been ordered differently or are they the same? If it is a combination we have 10 instead of 20 different mutations.

3. At mutation 3 we have 5*4*3 and at 4 we have 5*4*3*2 or 120 permutations.

4. At any location we may have any one or a combination of these mutation types. There are two factors driving their number:

a. A single type will have growth, dispersion and movement dynamics with the above mentioned model but each mutation will respond differently since their coefficients will be different. Some may grow faster and some may diffuse faster. There is no a priori ranking of the coefficients.

b. The surrounding mutant types will also tend to mitigate growth.

Now we can call the mutant cells as follows:

$n_{1,k}(x,t)$
$n_{2,j,k}(x,t)$
$n_{3,i,j,k}(x,t)$
$n_{4,h,i,j,k}(x,t)$
$n_{5,g,h,i,j,k}(x,t)$

Now we know how the subscripts can be ordered as per the above conversation. But we also assume that there exist some Markov mode for transitions from a 1 state to a 2 state, namely from one mutation to 2 mutations. That is we can assume a discrete time discrete state system and ascribe a Markov model with transition probabilities. Namely:

$$P\left[x(k+1)=x_0 \mid x(k)=x_0\right]=p_0$$
$$P\left[x(k+1)=x_{1,k} \mid x(k)=x_0\right]=p_{k,0}$$

Where we have the standard closure conditions on the transition probabilities. The process is Markov and it depends solely upon the prior state and no other.

In general we have:

$$P\left[x(k+1)=x_{n,k}\,|\,x(k)=x_{n,k}\right]=p_{n,k}$$
$$P\left[x(k+1)=x_{n+1,m}\,|\,x(k)=x_{n,k}\right]=p_{n:m,k}$$

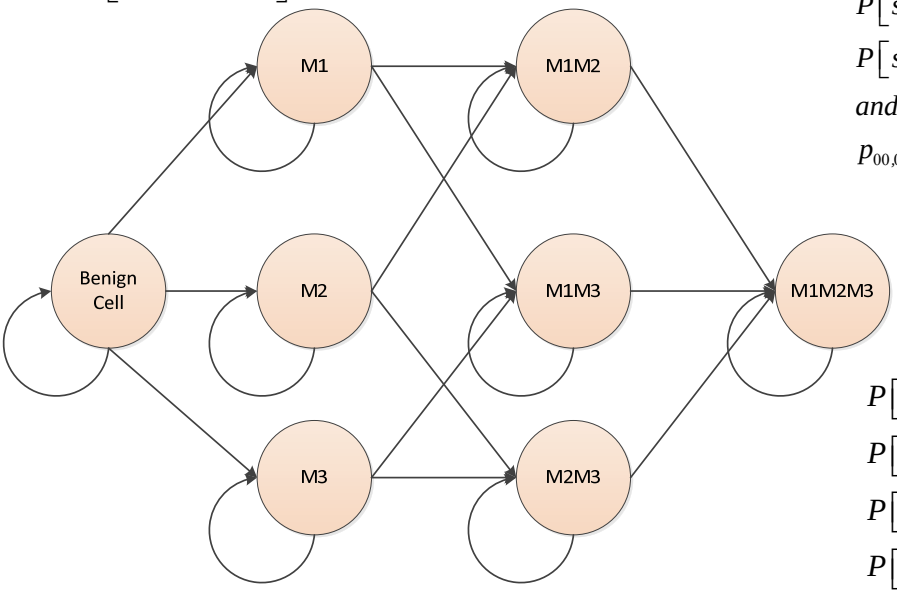
We can extend this to a continuous time system simply. We just need the mutation rates.

Now the initial equation describing single mutant growth was deterministic. However we now have demonstrated a random process. Thus we want to determine the average number at any time and at any location. The average will include the temporal Markovian dynamics of cell mutation rates which themselves may be spatially dependent.

8.2.2 The Markov Example

For example consider the following three gene mutation case:

$$P=\begin{bmatrix} p_{11}\cdots p_{21}\cdots p_{81} \\ \cdots \\ p_{18}\cdots p_{28}\cdots p_{88} \end{bmatrix}$$



$$P\left[s00|s00\right]=p_{00,00}$$
$$P\left[s11|s00\right]=p_{11,00}$$
$$P\left[s12|s00\right]=p_{12,00}$$
$$P\left[s13|s00\right]=p_{13,00}$$
and
$$p_{00,00}+p_{11,00}+p_{12,00}+p_{13,00}=1$$

$$P\left[s11|s11\right]=p_{11,11}$$
$$P\left[s21|s11\right]=p_{21,11}$$
$$P\left[s22|s11\right]=p_{22,11}$$
$$P\left[s23|s11\right]=p_{23,11}$$
and
$$p_{11,11}+p_{21,11}+p_{22,11}+p_{23,11}=1$$

Now for this simple example we can assume that 1 is the start and 8 is the end. We also assume that 2, 3, 4 are the first steps and 5, 6, 7 are the second. Then we have a transition probability matrix P as follows:

$$P = \begin{bmatrix} p_{1,1}..p_{2,1}..p_{2,2}..p_{2,3}..0.....0.....0.....0 \\ 0.....p_{2,2}..p_{2,5}..p_{2,6}..p_{2.7}..0.....0.....0 \\ 0.....0.....p_{3,3}..p_{3,5}..p_{3,6}..p_{3,7}..0.....0 \\ 0.....0....0.....p_{4,4}..p_{4,5}..p_{4,6}..p_{4,7}.....0 \\ 0.....0....0.....0.....p_{5,5}..0.....0.....p_{5,8} \\ 0.....0....0.....0.....0.....p_{6,8}.....0.....p_{6,8} \\ 0.....0....0.....0.....0.....0.....p_{7,8}....p_{7,8} \\ 0.....0....0.....0.....0.....0.....0.....p_{8,8} \end{bmatrix}$$

Where the sum across any row is unity. Thus we have:

$$p(k+1) = Pp(k)$$

where

$$p(k) = \begin{bmatrix} p_1(k) \\ ... \\ p_8(k) \end{bmatrix}$$

In this specific example.

Where p is the probability vector of all possible states, say 8 in our previous example, and P is the transition probability matrix. Note we have assigned P as dependent upon x and t. The same holds true for p.

Let us consider a simple example of N possible states and we examine the probability density of a specific state, Let:

$$p(n_i; x.t)$$

Where p is the probability density function of the random variable n on gene mix i. We assume that n is a continuous random variable so we can provide a simpler representation.

Now we can move to a continuous time simply by stating:

$$\frac{dp(x,t)}{dt} = P(x,t)p(x,t)$$

However we can simplify this as follows:

$$\frac{dp(n_i;x,t)}{dt} = -\lambda_{ii}p(n_i;x,t) + \sum_{j=1,j\neq i}^{N} \lambda_{i.j}p(n_j;x,t)$$

This is the continuous version of the above and is a simple Markov birth-death type model.

Now consider any point x, t in space-time. We can define:

$$E[n_i(x,t)] = \int n_i(x,t)p(n_i;x,t)dn_i$$
$$= \overline{n_i(x,t)}$$

We shall use this in the next section.

8.3 COMBINING

We now will combine the propagation equation model for a single malignant state with the overall evolution of malignant states via a Markov model.

Now consider a specific cell count at some specific x and t. We know we have:

$$\frac{\partial n_k(x,t)}{\partial t} = Ln_k(x,t)$$

Where L is the spatial operator.

We also have for the probability of each n being a specific value at a specific x and t as:

$$\frac{\partial p_k(x,t)}{\partial t} = \sum_{n=1}^{N} \lambda_n(x,t)p_n(x,t)$$

Now we can define the average of any n as follows:

$$\overline{n_k(x,t)} = \sum p_k(n_k(x,t) = n_m)n_m(x,t)$$

This is the average value of that specific n value for that specific gene mutation at that specific time and location. We have a means to calculate each of the probabilities and we then just calculate the average.

Now we can combine the equations as follows:

$$\sum p_k(x,t)\frac{\partial n_k(x,t)}{\partial t} = \sum \lambda_m(x,t)p_m(x,t)Ln_k(x,t)$$

or

$$\frac{\partial \sum p_k n_k}{\partial t} = L\sum \lambda_{m,k}(x,t)p_{m,k}(x,t)n_k(x,t)$$

or

$$\frac{\partial \overline{n_k(x,t)}}{\partial t} = L\lambda_k \overline{n_k(x,t)}$$

Let me redo the notation again. We will assume that n can be continuous for notation purposes. Thus we have:

$$\overline{n_k(x,t)} = \int p_k(n_k;x,t)n_k(x,t)dn_k$$

Now let us remember that for N possible n types we have:

$$\frac{dp_k(u_k;x,t)}{dt} = \sum_{j=1}^{N} \lambda_{k,j}p_j(u_j;x,t)$$

We have N equations like this.

If we multiply the propagation equation for any n as follows:

$$\frac{p_k(n_k;x,t)\partial n_k(x,t)}{\partial t} = p_k(n_k;x,t)Ln_k(x,t)$$

Now introduce the integral:

$$\int p_k(n_k;x,t)\frac{\partial n_k(x,t)}{\partial t}dn_k = \int\ p_k(n_k;x,t)Ln_k(x,t)dn_k$$

But recall that we can write:

$$\frac{\partial pn}{\partial t} = p\frac{\partial n}{\partial t} + n\frac{\partial p}{\partial t}$$

Then using what we know of p we have:

$$\frac{\partial p_k(n_k;x,t)}{\partial t} = -\lambda_{k,k}p_k(n_k;x,t) + \sum_{j=1;j\neq k}^{N} \lambda_{k,j}p_j(n_j;x,t)$$

Substituting and rearranging:

$$p_k(n_k;x,t)\frac{\partial n_k(x,t)}{\partial t}=\frac{\partial[\,p_k(n_k;x,t)n_k(x,t)]}{\partial t}-n_k(x,t)\frac{\partial p_k(n_k;x,t)}{\partial t}$$
$$=\frac{\partial[\,p_k(n_k;x,t)n_k(x,t)]}{\partial t}-n_k(x,t)\lambda_{k,k}p_k(n_k;x,t)+n_k(x,t)\sum_{j=1;j\neq k}^{N}\lambda_{k,j}p_j(n_j;x,t)$$

Now we can show that when we integrate to obtain averages we have:

$$\int p_k(n_k;x,t)\frac{\partial n_k(x,t)}{\partial t}dn_k=\frac{\partial\overline{n_k(x,t)}}{\partial t}-\lambda_{k,k}\overline{n_k(x,t)}-\sum_{j=1;j\neq k}^{N}\lambda_{k.j}n_j(x,t)$$

Thus we can replace this in the propagation equation to read:

$$\frac{\partial\overline{n_k(x,t)}}{\partial t}-\lambda_{k,k}\overline{n_k(x,t)}-\sum_{j=1;j\neq k}^{N}\lambda_{k.j}n_j(x,t)=L\overline{n_k(x,t)}$$

Now if we had multiplied by the joint pdf we would have obtained:

$$\frac{\partial\overline{n_k(x,t)}}{\partial t}-\lambda_{k,k}\overline{n_k(x,t)}-\sum_{j=1;j\neq k}^{N}\lambda_{k.j}\overline{n_j(x,t)}=L\overline{n_k(x,t)}$$

Or; upon rearranging we obtain:

$$\frac{\partial\overline{n_k(x,t)}}{\partial t}=L_k\overline{n_k(x,t)}-\lambda_{k,k}\overline{n_k(x,t)}+\sum_{j=1;j\neq k}^{N}\lambda_{k.j}\overline{n_j(x,t)}$$

Thus the result for the average is a set of linked partial differential equations. Note we have modified the L operator to reflect specificity for k. The added terms reflect the movement of cell types from one class to another.

This is a powerful equation. It tells us how specific cells diffuse, flow and reproduce, and then how they migrate to new types of cells.

Let us take it one step further. Recall:

$$L_k=a_k\frac{\partial^2}{\partial x^2}+b_k\frac{\partial}{\partial x}+c_k$$

define

$$\widetilde{L_k}=L_k-\lambda_{k,k}$$

Now consider a vector of all n possibilities and we can determine the average vector of these as follows:

$$n(x,t) = \begin{bmatrix} n_1(x,t) \\ \dots \\ n_N(x,t) \end{bmatrix}$$

And where the average of the vector is the average of the above. Then we readily have the equation for all n as follows:

$$\boxed{\frac{\partial \overline{n(x,t)}}{\partial t} = \tilde{L}\overline{n(x,t)} + \Lambda \overline{n(x,t)}}$$

Where:

$$\tilde{L} = \left[\tilde{L}_1 \dots . \widetilde{L_N} \right]$$

and

$$\Lambda = \begin{pmatrix} -\lambda_{11} & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & -\lambda_{22} & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & -\lambda_{33} \end{pmatrix}$$

The above is suggestive and it depends on the specific model.

8.4 A PHENOMENOLOGICAL MODEL FOR SURVIVAL ANALYSIS

The use of statistics had developed extensively in the early 20th Century, Methods had been applied to such issues as Hypothesis testing, namely were two results the same or were that not the same. Thus came the idea of a null hypothesis. In the area of survival analysis one saw this applied to Medicine using the Kaplan Meir analysis, a method to estimate a survival curve, and then examining the way to ascertain whether two survival curves were essentially the same or different; a variant on the null hypothesis.

The Statistician approach was not to assume any underlying physical or phenomenological model, just look at the data and make gross statistical statements

Then came in the 1960s the models using underlying physical systems and estimators which gave estimates and variances. These were the Kalman filters which in a sense were offshoots of the work of Wiener and Kolmogorov in the 1940s. The basis for their work all assumed some underlying physical reality and a desire to estimate the value of some measure of that reality as time proceeds.

Following from that was the development of Identification methods, wherein one iteratively looked at reality and then sought to infer, estimate, the parameters defining that reality.

Now in this analysis we re-examine survival analysis but in a manner reflective of some underlying reality. Survival curves can be used, if we include that reality, as a way to estimate or identify the specific differences between the two disease states.

Let us begin with a simple example. We assume we have a simple death process. Thus the percent of the initial population at any time is given by the following temporal dynamics:

$$\frac{dp(t)}{dt} = -\mu p(t)$$

or

$$p(t) = \exp(-\mu t)$$

Here we have an instantaneous death rate μ.

In the more complex models we often have the following:

$$p(t) = \exp(-H(t))$$

where H is some complex time varying number which we desire to estimate and compare. Yet in the classic case there is no phenomenological connection between the H(t) and anything happening in the disease process.

Let us now consider how one might add a phenomenological element.

We have demonstrated elsewhere that one can build a model for the number of malignant cells as a function of space and time. That model is:

Propagation Model: This equation provides a spatio-temporal model for the calculation of the number of specific cancer cells which are propagated by means of: (i) diffusion, (ii) flow, and (iii) proliferation.

$$\frac{\partial n(x,t)}{\partial t} = a\frac{\partial^2 n(x,t)}{\partial x^2} + b\frac{\partial n(x,t)}{\partial x} + cn(x,t)$$

Average Model: This model considers the calculation of the average number of malignant cells in a spatio-temporal manner when the cells mutate into N possible genetic variants. It calculates the average number by variant and thus is a vector equation containing the N variants.

$$\frac{\partial \overline{n(x,t)}}{\partial t} = \tilde{L}\overline{n(x,t)} + \Lambda\overline{n(x,t)}$$

Now how does this relate to the survival function? Simply. If the total number of cancer cells exceeds a certain number then the patient dies. The summary is below. Let us first define the metric:

$$N_{Tot}(t) = \int \overline{n(x,t)}dx = N$$

Thus we have the following:

State	*Number of Cells: N*	*Status*
Localized	10^8	Visible with Image Contrast
Disseminating	10^9	Palpable
Metastatic	10^{12}	Death

Thus we can relate the non-phenomenological model of H(t) to one wherein we take advantage of what we now can know about the cell. That is:

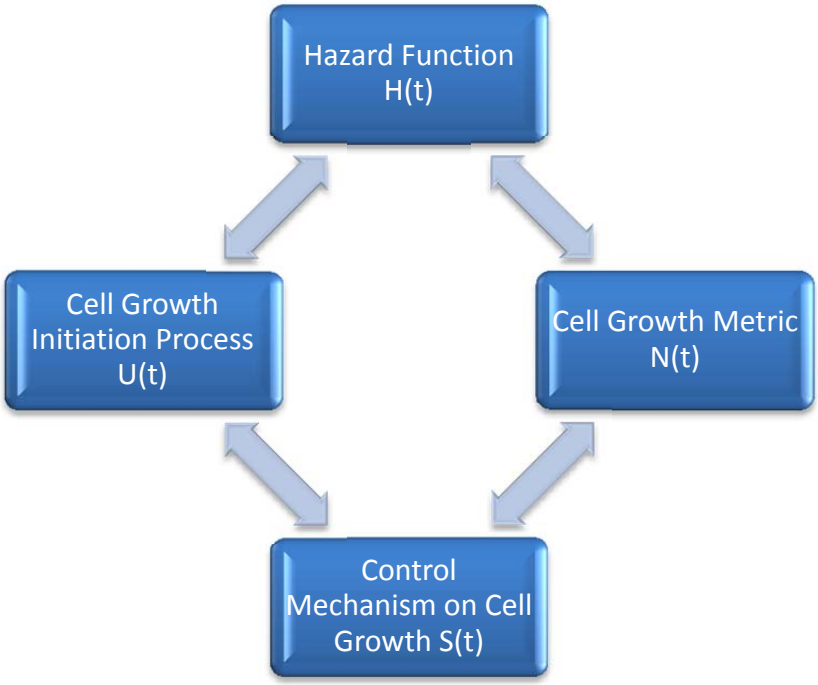
$$H(t) = h(N_{Tot}(t))$$

We cannot say yet just what that relationship is. But unlike all other survival models we have a nexus between survival and the disease. For example we can hypothesize:

$$H(t) = kN(t)$$

which is a standard linear relationship? Note that N(t) is growing exponentially and thus we have an exponentially enlarging H(t) and depending on how fast will determine the presentation of the Kaplan Meier curve data.

In turn we have a nexus between the disease and treatment which we demonstrate below.



Notice that we have added two additional dimensions. We have a relationship between H(t) and N(t) and we also can relate N(t) to some variable we call U(t) the underlying dynamics of cell mutation and growth, and also to S(t) the cell control mechanisms which can represent the therapeutic effects.

For example if we have a BRAF V600 mutation than that is a U(t) factor and the inhibitor drug is the S(t) factor.

8.5 OBSERVATIONS

To summarize the following depicts the major analytical results:

Propagation Model: This equation provides a spatio-temporal model for the calculation of the number of specific cancer cells which are propagated by means of: (i) diffusion, (ii) flow, and (iii) proliferation.

$$\frac{\partial n(x,t)}{\partial t} = a\frac{\partial^2 n(x,t)}{\partial x^2} + b\frac{\partial n(x,t)}{\partial x} + cn(x,t)$$

Average Model: This model considers the calculation of the average number of malignant cells in a spatio-temporal manner when the cells mutate into N possible genetic variants. It calculates the average number by variant and thus is a vector equation containing the N variants.

$$\frac{\partial \overline{n(x,t)}}{\partial t} = \tilde{L}\overline{n(x,t)} + \Lambda\overline{n(x,t)}$$

We can thus make several important observations regarding this model.

1. Prognostic and Therapeutic: We can determine the transitions and the factors related to diffusion, flow and growth. Thus we can use the result as a powerful one for prognostic and therapeutic results. As we had indicated earlier, the Observability and Controllability issues are essentially Prognostic and Therapeutic respectively.

2. Variances: The results are for the average. We can determine the results for the variances as well. We have examined the variances on the averages and they are somewhat complex and we do not believe that they lend significant additional information at this time.

3. Solutions: The solutions to these equations are readily obtained using standard techniques. They can, in addition, be determined in closed form results.

8.6 MEASURING THE PARAMETERS

Measuring the parameters in these models has been discussed before for a single mutation. However, we have now introduced a set of multiple and progressive mutations. How do we know how these mutations progress? Can the mutation progressions be determined a priori or do they occur in some random fashion? How do we deal with the epigenetic elements such as hypermethylation and miRNAs when we consider changes in expression without mutations?

There are a significant number of questions that we must consider when examining the change in gene expression. In addition the environment, extracellular matrix as well as surrounding cells may also effect changes in gene expression.

Thus ascertaining gene mutation or expression would most likely be determined by examining the cells expression themselves. At the current time there does not appear to be a robust theory which can be used as a basis for such projections. We are left measuring what has actually happened rather than projecting what will occur.

Perhaps subsequent Bayesian analysis will allow for such determination.

8.7 IN SITU HEMATOLOGICAL MEASUREMENTS

The challenge is determining of a cancer has metastasized is to find out where and how much. The classic approach is to look at the local draining lymph nodes and see if has gone there. However the cancer cells may often escape through the blood system and not the lymph system. Consider ocular melanoma, there is no lymph system connection and it spreads by hematological means only.

That means that by examining the blood we should be able to find the wandering malignant cells, at least in theory. In a recent release by *MedGadget* the article relates developments at MGH in Boston as follows[5]:

Circulating tumor cells (CTCs) are shed by primary tumors and allow the cancer to metastasize to the distant sites. While this is a devastating tool in cancer's war chest, it offers clinicians a marker through which to diagnose and monitor progress of the disease. Since the discovery of CTCs over a hundred years ago, researchers have been developing ever more sensitive methods of capturing them since they're extremely rare in whole blood.

In a recent development by Ozkumur et al at MGH[6] the authors' state:

Circulating tumor cells (CTCs) are shed into the bloodstream from primary and metastatic tumor deposits. Their isolation and analysis hold great promise for the early detection of invasive cancer and the management of advanced disease, but technological hurdles have limited their broad clinical utility. We describe an inertial focusing–enhanced microfluidic CTC capture platform, termed "CTC-iChip," that is capable of sorting rare CTCs from whole blood at 107 cells/s.

Most importantly, the iChip is capable of isolating CTCs using strategies that are either dependent or independent of tumor membrane epitopes, and thus applicable to virtually all cancers. We specifically demonstrate the use of the iChip in an expanded set of both epithelial and nonepithelial cancers including lung, prostate, pancreas, breast, and melanoma.

The sorting of CTCs as unfixed cells in solution allows for the application of high-quality clinically standardized morphological and immunohistochemical analyses, as well as RNA-based single-cell molecular characterization. The combination of an unbiased, broadly applicable, high-throughput, and automatable rare cell sorting technology with generally accepted molecular assays and cytology standards will enable the integration of CTC-based diagnostics into the clinical management of cancer.

There are several problems here however:

1. As we had demonstrated in some of our prior analysis, blood borne cancer cells are rare, but more importantly they are cells which are coming from and going to organs. Namely they are in transit, from whence and to where we do not know.

2. The genetic states of each of these wandering cells may be a marker of from whence it came. The problem is that we do not fully understand this genetic mutation process, and in fact as we have shown before it may actually be a Markov like chain process.

[5] http://www.medgadget.com/2013/04/mgh-ctc-ichip-sets-new-bar-for-circulating-tumor-cell-detection.html

[6] http://stm.sciencemag.org/content/5/179/179ra47

3. Understanding this change in cells may be of significant therapeutic value. However this again is uncertain given our current state of knowledge.

4. Again we come back to the cancer stem cell and ask if the few cells we find in the blood stream are the right cells to examine.

However this advance could provide significant data to allow us to expand the understanding of mutating cancer cells.

It seems that there is a significant amount of new work being done on evaluating cancers via circulating tumor cells and their DNA. Another paper in Nature states:

Cancers acquire resistance to systemic treatment as a result of clonal evolution and selection. Repeat biopsies to study genomic evolution as a result of therapy are difficult, invasive and may be confounded by intra-tumour heterogeneity Recent studies have shown that genomic alterations in solid cancers can be characterized by massively parallel sequencing of circulating cell-free tumour DNA released from cancer cells into plasma, representing a non-invasive liquid biopsy.

Here we report sequencing of cancer exomes in serial plasma samples to track genomic evolution of metastatic cancers in response to therapy. Six patients with advanced breast, ovarian and lung cancers were followed over 1–2 years. For each case, exome sequencing was performed on 2–5 plasma samples (19 in total) spanning multiple courses of treatment, at selected time points when the allele fraction of tumour mutations in plasma was high, allowing improved sensitivity.

For two cases, synchronous biopsies were also analysed, confirming genome-wide representation of the tumour genome in plasma. Quantification of allele fractions in plasma identified increased representation of mutant alleles in association with emergence of therapy resistance. ...treatment with gefitinib.

These results establish proof of principle that exome-wide analysis of circulating tumour DNA could complement current invasive biopsy approaches to identify mutations associated with acquired drug resistance in advanced cancers. Serial analysis of cancer genomes in plasma constitutes a new paradigm for the study of clonal evolution in human cancers.

Cancer Research UK commented on the works as follows[7]:

Scientists ... used traces of tumour DNA, known as circulating tumour DNA (ctDNA) found in cancer patients' blood to follow the progress of the disease as it changed over time and developed resistance to chemotherapy treatments.

[7] http://www.cancerresearchuk.org/cancer-info/news/archive/pressrelease/2013-04-07-simple-blood-test-to-track-tumour-evolution?rss=true

They followed six patients with advanced breast, ovarian and lung cancers and took blood samples, which contained small amounts of tumour ctDNA, over one to two years.

By looking for changes in the tumour ctDNA before and after each course of treatment, they were able to identify which changes in the tumour's DNA were linked to drug resistance following each treatment session.

Using this new method they were able to identify several changes linked to drug-resistance in response to chemotherapy drugs such as paclitaxel (taxol) which is used to treat ovarian, breast and lung cancers, tamoxifen which is used to treat oestrogen-positive breast cancers and transtuzumab (Herceptin) which is used to treat HER2 positive breast cancers.

And they hope this will help shed new light on how cancer tumours develop resistance to some of our most effective chemotherapy drugs as well as providing an alternative to current methods of collecting tumour DNA – by taking a sample direct from the tumour – a much more difficult and invasive procedure.

As we noted in a previous note regarding the same set of procedures by others researchers this is a useful method to detect the progression of cancer.

However the following observations are of note:

1. Are these coming or going cells, namely are the cells on their way to a metastasis or the result of one.

2. Can we use these cells to determine the changes in DNA expression as the cells progress.

3. How effective a prognostic tool are these measurements.

4. What therapeutic methods can be applied now knowing this information.

Thus is this data of primary use or secondary. Notwithstanding its clinical use it does represent an excellent tool for genomic progression.

8.8 NON INVASIVE METHODOLOGIES

Molecular Functional Imaging, MFI, provides a set of non-invasive methodologies to ascertain pathway dynamics as well as changes in genetic expression[8]. As Glunde et al state:

Molecular–functional imaging (MFI) can be defined as the noninvasive visualization of molecular and functional pathways in the tissue of interest.

As Glunde et al state:

[8] Further consideration of the material in this section was provided by discussions with Dr. Annick Van den Abbeele at Dana Farber Cancer Institute, Boston, MA.

Molecular imaging of cancer detects single molecules or their activity within cancer cells in culture or within a solid tumor. These molecules can be overexpressed receptors, activated enzymes or relocated molecules, each of which plays important roles in signaling cascades or regulatory programs that are deregulated in cancer. These in turn give cancer its phenotypic characteristics, such as evasion of apoptosis, self-sufficiency in growth signals, insensitivity to anti-growth signals, sustained angiogenesis, unlimited replicative potential, invasion of tissue and metastasis.

They continue focusing on the ability with MFI to monitor pathway dynamics:

The ability to image gene expression, promoter activity, and transcriptional activity in vivo is important because these are the starting points for many deregulated pathways in cancer. Reporter genes that are typically used include luciferase genes for bioluminescence imaging, fluorescent-protein genes for fluorescence imaging, herpes simplex virus thymidine kinase (HSVtk) genes for PET and SPECT imaging and ferritin genes or chemical exchange saturation transfer (CEST) reporters for MRI. These reporter genes are placed under the control of a promoter of interest so that promoter activity in vivo can be evaluated.

Triple-fusion-reporter genes that allow for in vivo multi-modality imaging with bioluminescence, fluorescence and PET have recently been developed. Imaging gene expression has helped to delineate mechanistic and functional aspects of oncogenes, such as myc, and tumor suppressor genes, such as p53. Another important application of imaging gene expression is monitoring viral vector delivery in vivo for future gene therapies.

8.8.1 Pathway Expression Estimation

Glunde et al discuss two specific genes; myc and p53. Let us first consider what they say about myc:

The myc oncogene is one of the most commonly activated oncogenes associated with the pathogenesis of liver cancer. In vivo bioluminescence imaging of transgenic mice conditionally expressing tetracycline-inducible myc proto-oncogene in liver cells proved that myc oncogene inactivation resulted in dormancy as long as myc remained inactive but that myc reactivation immediately restored the neoplastic features of previously differentiated hepatocytes and biliary cells. In this study, myc activation probably caused malignant expansion of immature liver cells with stem-cell like properties, supporting the hypothesis that liver tumors can arise from cancer stem cells.

The cancer stem cell issue is also a significant one in this analysis. We have discussed cancer stem cells previously and they are key factors in assessing metastatic behavior. However they are also of low density and generally difficult to describe genetically.

Glunde et al then progress to a discussion of p53:

The p53 tumor-suppressor gene, which is mutated in rv50% of all human cancers, plays a key role in cell cycle regulation and apoptosis following DNA damage by functioning as a sequence-specific transcription factor. Bioluminescence imaging was employed for the noninvasive evaluation of the transcriptional activity of p53 in vivo in a transgenic mouse model that conditionally expressed the firefly luciferase gene upon activation by a p53-responsive promoter.

After exposure to ionizing radiation, the in vivo p53 transcriptional activity displayed a distinct oscillatory pattern, confirming p53 transcriptional oscillations previously observed in cultured cells. In vivo bioluminescence imaging of transgenic mice will prove useful in future studies assessing p53 response in vivo after systemic administration of novel therapeutic p53 or proteasome inhibitors or of agents modulating the response to ionizing radiation.

The following Table is an adaptation from the paper summarizing some targets for MFI:

Cancer characteristics/pathways	Molecular/cellular targets in cancer	Molecular MFI applications
Oncogenesis pathways	p53	Bioluminescence, PET
	myc	Bioluminescence
Multidrug resistance	P-glycoprotein	PET, SPECT
Apoptosis	Phosphatidyl serine	PET, SPECT, MRI, Optical
	externalization	(Annexin V)
Cell surface receptors	EGFR	PET, SPECT, NIR fluorescence
	HER-2/neu	T1-MRI, T2-MRI, PET, SPECT,
		fluorescence
	PSMA	PET, SPECT, fluorescence
Proliferation/differentiation	Thymidine kinase	PET (^{18}F-FLT)
	Telomerase	PET
Angiogenesis/lymph-	VEGF	PET, SPECT, MRI, fluorescence
angiogenesis		
Hypoxia	avb3 HIF-1	Fluorescence, PET, SPECT, MRI PET, fluorescence
Metabolism	NA	NA
ECM degradation	Cathepsin D	NIR fluorescence
	Cathepsin B	NIR fluorescence
	Matrix	NIR fluorescence
	metalloprotease 2	
	Lysosomes	Fluorescence
Invasion and metastasis	Cell labeling with	Fluorescence
	fluorescent proteins	

8.8.2 ECM Imaging

We have discussed the impact of the extracellular environment for melanoma in previous sections and thus being able to deal with that is critical as well. Thus as Glunde et al sate:

A growing list of imaging techniques such as differential interference contrast (DIC) microscopy, confocal reflection microscopy, second harmonic generation (SHG) microscopy, Fourier transform infrared (FTIR) micro-spectroscopy and atomic-force microscopy (AFM) (reviewed in) are helping us to understand the interaction be- tween tumor cells and the ECM. For example, Nomarski DIC optics has been used for dynamically tracking cell- induced matrix remodeling. Assembly properties of type I collagen and interstitial ECM have been studied with confocal reflection microscopy, without the need for staining the specimen.

The nonlinear optical process of SHG requires an environment without a center of symmetry, such as an interfacial region, to produce a signal. SHG was proposed as a new contrast mechanism for live- cell imaging and was extended to image endogenous structural proteins such as those within collagen-rich layers. FTIR micro-spectroscopy, which probes the vibration energy of chemical bonds, has been used for imaging the proteolytic activity of matrix metalloproteinases (MMPs) produced by invasive cancer cells on collagen-based matrices

8.8.3 Next Steps

Glunde et al list the following next steps for MFI:

- *Further development of novel strategies to detect and image specific pathways and targets is required.*
- *The low concentration of receptors and molecular targets and the inherent insensitivity of imaging techniques, such as MRI, impose limitations, which require amplification strategies for increasing the sensitivity of detection.*
- *Intracellular access of reporter molecules or the internalization of reporters is necessary for imaging several critical pathways and molecules and presents another challenge.*
- *As the acquisition of multi-parametric and multi-modality images becomes increasingly routine, integrated multi-modal approaches will present unique technical and computational challenges as well as exciting opportunities.*
- *The limited translation of these approaches to clinical use presents the major challenge to date.*

The summary above clearly indicates a significant potential if the challenges are met but they also present a significant set of current limitations.

With molecular functional imaging it is possible to do the following:

1. Target specific pathway modalities so as to ascertain the state of the cells in any part of the body and to determine the mutation profile of the malignancy. Specifically we can determine the progression from the initial mutation to subsequent ones. It is suspected that mutation profiles will be of great assistance in both prognostic as well as therapeutic approaches.

2. Spatial progression will be able to be identified, quantified and projected. True assessment of metastatic movement can be made at the molecular cell level.

3. Temporal changes will be the most critical of measurements. We will be able to examine rates of change both in metastasis as well as in response to therapeutic advances.

MFI can provide a critically impressive capability which when combined with the modeling described herein will allow for improved prognostic capabilities as well as improved therapeutics.

9 CONCLUSIONS

We have developed a methodology to determine the growth and spread of cancer cells as they change their genetic expression, internally and externally. Like any model the one developed herein has strengths and weaknesses but it does present a predictive method for analyzing cancers. It also is a method which looks at cancer on a macro level yet retaining the fundamental elements of gene expression and protein operation. It allows for the inclusion of epigenetic factors which is exceptional in most models. The very nature of its predictive use is a distinguishing characteristic.

The recent report on such a cancer prognostic model such as Oncotype DX by Knezvic et al is a putatively prognostic method used in prostate cancer. Fundamentally what they do is examine cancer cells for the expression of various genes and examine three sets; baseline expressions, excess expressions and reduced expression. They use the baseline to set levels for excess and reduced. They then use the excess or reduced in a one dimension expression to determine a prognostic measure. Like PSA measures, CA125, CEA, and the like, they try to reduce everything to a single number. We argue here that such an approach is problematic at best. Furthermore they fail totally to demonstrate any internal pathway influence. There is no predictive basis for their approach predicated upon the actual dynamics of the cell. It is purely correlative and there may be substantial confounders involved. This approach is an example of what we fell to be the poorer aspects of genomics applied to cancer prognostics.

9.1 APPROACH

The approach which we have developed herein is on the one hand an extension of other approaches but on the other hand a novel jump to establishing a body level of a measurable and predictable expression of cancer cell loading. We use a tempero spatial model acceptable in many domains and then augment it with a genetic state model wherein we can incorporate the genetic changes that cause the aggressive development and spread of cancer. The combinations lead to the simple result we obtain in our final result.

The second most important contribution is the ability using the model to determine state coefficients via standard methods of system identification using now standard Lab bench techniques. We have demonstrated some specific applications thereof.

Thus we have developed a complete methodology for prognostic modelling and a method which can iteratively use data from experimental results.

Moreover, we have an underlying model of what should be against which we can compare what is. The iterative improvements will allow substantial improvements to the technique.

9.2 PRINCIPLES

Dougherty in his many papers has addressed this issue in considerable detail, namely he looks at what he calls the epistemological crisis in genomics[9]. Dougherty has written a brilliant set of papers which set the path for those in genomics but it should also be a warning for economists.

Let us look at Dougherty from the aspect of the genomist. Keep in the back of our minds the same issues for the economist.

Dougherty begins his paper by stating:

There is an epistemological crisis in genomics. At issue is what constitutes scientific knowledge in genomic science, or systems biology in general. Does this crisis require a new perspective on knowledge heretofore absent from science or is it merely a matter of interpreting new scientific developments in an existing epistemological framework? This paper discusses the manner in which the experimental method, as developed and understood over recent centuries, leads naturally to a scientific epistemology grounded in an experimental-mathematical duality.

He continues:

The change brought about by the "new science" of the Sixteenth and Seventeenth Centuries is based on the integration of two principles: (1) design of experiments under constrained circumstances to extract specifically desired information; and (2) mathematical formulation of knowledge.

The two principles arise from the two sides of the scientific problem, the source of knowledge and the representation of knowledge in the knower. Perhaps the greater revolution in knowledge is the design of experiments. One need only think of Archimedes' mathematical analyses of fluidics and mechanics to see that the ancients recognized the central role of mathematics, even if they did not understand that role in the modern sense. But the modern concept of experiment is a different matter altogether.

Dougherty then proceeds to emphasize the need and usefulness of the mathematical model, namely the ability to predict. He specifically states:

A mathematical model alone does not constitute a scientific theory. The model must be predictive. Mathematics is intrinsic because science is grounded in measurements; however, a model's formal structure must lead to experimental predictions in the sense that there are relations between model variables and observable phenomena such that experimental observations are in accord with the predicted values of corresponding variables.

These predictive relations characterize model validity and are necessary for the existence of scientific knowledge.

The predictive ability of a model is essential. Frankly the ability to go from experiments to prediction, albeit with uncertain but projective results, is at the heart of a scientific approach and

[9] Dougherty, On the Epistemological Crisis in Genomics.

it is essential to the engineering approach that we aim for herein. Just having the experimental data is necessary but not sufficient. One must be able to link the data in some mathematical models which in turn is in itself predictive of alternative and future results. Dougherty continues:

The fundamental requirement of a scientific validation procedure is that it must be predictive. A scientific theory is not complete without the specification of achievable measurements that can be compared to predictions derived from the conceptual theory. Moreover, it depends on the choice of validity criteria and the mathematical properties of those criteria as applied in different circumstances.

Again in genomics we have collections of data, and as we shall demonstrate herein for prostate cancer, that data is enlightening and descriptive but is not predictive and the mathematical models of what is transpiring is missing. It is not that we cannot construct a model, and indeed such a construct would itself be subject to continuing change as we learn more, but the intent must be there to construct such a model, and a framework, a paradigm, must be ever present.

Dougherty makes another interesting observation as follows:

Consider the following statement of Steven Jay Gould: "Science tries to document the factual character of the natural world, and to develop theories that coordinate and explain these facts". Perhaps this statement would have been accurate during medieval times, but not today. While it is true that theories coordinate measurements (facts), it is not the documented measurements that are crucial, but rather the yet to be obtained measurements. Gould's statement is prima fascia off the mark because it does not mention prediction.

The point that Dougherty is making is that science, the science of systems particularly, which we will discuss herein, focuses on not the facts, but the process, and the predictability of the models we develop from facts, connected and ordered facts, facts which are quantified. In Gould's world view, that of a botanist or zoologist of the early 20th century, he is a collector and organizer of facts. He is not the builder of models which can be used to predict forward based on prior data. True systems approaches are valid if and only if they have the predictability capability.

As Dougherty states after this, the true nature of systems science and engineering is not the fitting of data, it is not the clustering of data points in some microarray. As Dougherty quotes Norbert Wiener from Wiener's classic, Cybernetics, the true nature of systems is a combination of models, predictability and uncertainty.

Finally Dougherty comments on the use and abuse of logistic analysis. Simply put he demonstrates via the arguments of others, especially William Feller, that frankly one can get a logistic fit to anything. It does not demonstrate any physical causality. In physics we have the relationship between mass and momentum, through the variable velocity.

We have rate equations for reactions demonstrable by experiment and validatable by predictive use. We understand forces and torques so that we can analyze and design a bridge, and we understand the multi body problem so that we can predict what trajectory a rocket may take as it

traverses the galaxy. For genomic analysis and especially for cancer can we develop such a methodology.

9.3 EPISTEMOLOGY AND MODELS

The book, ***Epistemology of the Cell***, by Dougherty and Bittner is a gem. Now, systems-thinking was, in way, started by Norbert Wiener in the 1930s, as he began to model various systems, from the dynamics of nerve fibers to the development of the first artificial arm. Yes, he did the arm, not the physician at Mass General. Wiener also developed the systems to the control the pointing of radar controlled guns on ships in WW II.

What made Wiener unique was that Wiener asked "Why". The why meant finding the cause and the result and establishing the connection between the two, establishing the system. Physicists and chemists ask "why" questions, the find the cause and the causality chain. Biologists were for ages asking "what" as people who classified.

Darwin broke that chain of biologists somewhat by asking "why" as regards to evolutions and he is pilloried even to this day. Physicians often ask "what" and "how" and do not really want to be bothered by the "why". It is not their jobs to find out why, just find out "what" is wrong and know "how" to fix it. Thus a physician is taught "what" to look for to diagnose prostate cancer and is also taught "how" to remove it. The physician does not know or care "why" the cancer is doing what it is doing. That is epistemological.

That, in essence, is a simplification of what Dougherty and Bitter go about to explain. Their book explores the challenges set forth to those exploring the gene and the cell and entices them to think beyond just the "what and how" to go to "the why", the explanation of the process, the system, from cause to effect. The authors have written a treatise which compares a few others which look at the epistemological basis of research, asking the correct questions, and pursuing the best path to answer them (see the work by Winograd and Flores as a prime example written some 25 years ago). Dougherty has with his colleagues and associate been developing the ideas contained herein for well over the past decade and I have read many of his works, they provide great insight to what should be done.

Dougherty begins with a discussion of the definition of epistemology[10]. They define scientific epistemology by what it addresses; knowledge and its truth. Then they use Kant to develop the transition from the Enlightenment to today. Then there is a brief but focusing discussion where he explains the simple Newtonian world maturing into an Einstein one and likewise a Watson and Crick paradigm of DNA/RNA/proteins into the way we understand cell dynamics today, as pathways, miRNAs, repressor enzymes and the complexity of both intracellular dynamics and intercellular dynamics[11]. They rely extensively on a Popperian view of Science, which for those more familiar with Kuhn may tend to have some slight dissonance, but it holds together quite well.

[10] Dougherty, p. 2.

[11] Dougherty, p. 4.

He sets the stage well, both from establishing the necessity to have models which answer why, and to establish the counterpoint of the thinkers who have reverted back to Aristotelian classification as the finders of what.

There is a discussion of Aristotelian causality where he states it well with the statement[12],

"explanation must involve a causal relation…".

They state, "Galileo and Newton do not deny causality as a category of knowledge but they widen the scope of knowledge to include mathematical systems that relate phenomena, while bracketing "questions about nature" of the phenomenon…"[13].

They have an excellent discussion of Bertrand Russell's work on causality[14].

The authors use the work of Norbert Wiener and his associate Arturo Rosenbluth on Cybernetics[15]. For it was indeed "the synergy of communications, control, and statistical mechanics…" that set the framework of how we should view cellular and organ dynamics. The authors then gave examples of gene regulation. I would have simply stated that every cell and every organic system is a multidimensional distributed random process.

One could take a Feynman like approach and posit the obvious, and then fill in the details. The authors work from the bottom up to demonstrate their world view. Namely that when we look at cells we are looking at complex dynamic random systems. Systems we can ascribe states to, states being measurable quantities, which in turn operate on other states in a dynamic fashion.

The authors start the transition to complex state models. They again rely on the wisdom of Wiener to state[16]:

"Wiener recognized the difficulties that the mathematical requirement of science and translational science would present for medicine …"

For back in 1948 when Wiener had published Cybernetics, Medicine was still a "what and how" practice. It did not transition to a "why" approach. The translational science that the authors speak of is:

"… mathematical engineering, applied mathematics with a translational purpose.."

[12] Dougherty, p 14.

[13] Dougherty, p 21.

[14] Dougherty, pp 34-35.

[15] Dougherty, p 70.

[16] Dougherty p 89.

Namely, to translate nature to measurable quantities. Quantities which we can then by knowing the "system" we can then manipulate and predict. We can observe and we can control, the end goals of translational science. In a Popperian sense, the authors address the issue of measuring, predicting, and examining what does not do what we said it would.

The key arguments are developed in the "sola fides" discussion, faith alone, as a mantra to those who fail to understand the system nature of the cell dynamics. The authors evoke William Barrett, the insightful Columbia University philosopher, who wrote The Illusion of Technique, a superb work integrating the principles of epistemology and science in the late 1970s[17]. Frankly, to see Barrett in a book of this type was an exciting surprise, for I had thought that Barrett was falling into obscurity, a loss to many who are struggling with issues that Barrett has thought through decades ago. The authors then discuss the nature of stochastic dynamic systems.

Dougherty brings insight via avenues that I found resonated strongly. The discussion on Wiener, where Dougherty, unlike Gleick in what I feel is presented with uninformed bias[18], sees Wiener as the father figure, one combining systems thinking, clear and built upon strong mathematical foundations, which is then integrated with real biological systems. Although I find their approach insightful and compelling, I would have taken pathways in cancer dynamics as somewhat well-defined stochastic systems.

For example, we know the effects of PTEN, the AKT pathway, and the MYC pathway, the p53 pathway, and the complex dynamics which are well described in the readily available NCI data base of pathways. One can use as states the concentrations of any one of these proteins and then state simply that they all interact with one another, the result being homeostasis or if a change cancer. The model is multidimensional, stochastic, highly complex, and strewn with "noise", namely uncertainties. Models have been developed and tested for such cancers as prostate, melanoma and colon. Dougherty, himself, has made substantial contributions to this area. It would have been useful perhaps to demonstrate this approach as well.

The clear strike at the adversaries is set. After again referring to Barrett and Kant, the authors end with[19]:

"Does anyone really believe that data mining could produce the general theory of relativity?"

I think this can be extended. For example many researchers run millions of microarrays and are currently finding hundreds of SNPs or thousands of miRNAs and each time they send out a press release saying they have "discovered" some new "gene" or worse "cause" or "cure" of say

[17] Dougherty pp 148-149.

[18] See Gleick, The Information, Vintage (New York) 2012. Gleick had written a book in praise of Shannon and deprecating Wiener. Gleick was in my opinion influenced by the Bell Labs wing of the debate and in fact it was Wiener and not Shannon who developed the basic principles.

[19] Dougherty p 165.

prostate cancer or melanoma. In reality one does not know whether this is a marker for cancer, a marker for a predisposition for cancer or just plain noise.

What the authors, and others, have argued is that it is essential to have a well-defined dynamic system model of how say PTEN and AKT interact and how they in turn control MYC and where the controls on p53 are in this chain. The microarray analyses should be done in the context of defining the linkages in the state model and not as ends in themselves. The model can then be validated. From such a model we can then see conditions on their way to cancer and conditions representing advanced cancers. For example, recent authors have announced a way to measure PTEN in prostate cancer and laud that as a diagnostic step. In reality by the time PTEN has been deactivated there is most likely a metastasis. Understanding and refining a model is the essence of the "why" articulated by the authors.

There is a superb critique of what the authors call the pre-Galilean thinkers, namely the biologists who like Linnaeus were really just classifiers of forms and shapes failing to understand why they were what they were[20]. One must remember that biology was all too often just a study of things and a process of naming them and classifying them. The systems which made for these differences were little understood, and worse, beyond the mindset of many who practiced in the field.

Chapter 9 is, in my opinion, the pinnacle of their argument. Simply put, we should now begin to perform our experiments within the context of a model. For example, we know many of the pathways of the key genes intracellular, but we do not yet understand the dynamic model that controls them. Thus, when we do microarray tests, we should be doing them to determine the constants in the model and then validate that design. We should, in effect, identify the system, using a system framework, not just some unstructured set of classifiers. We have the structures, now is the time to put them to use.

Science is iterative. It is an iterative set of models and refinements. The authors refer to Turing's last paper on tessellation, or why zebras have stripes[21]. The paper by Turing was submitted a day or two before he committed suicide and it was done without the benefit of Watson and Crick who were simultaneously doing their work at Cambridge. He intuited intercellular flow of some yet to be defined controlling substances. The concentration of these unknown substances would rise and fall in concentration and as such the color would change.

This approach has recently been applied, using a system model of flower genetics, and it explains and demonstrates the control of patterning in a genus of flowers. Having the model for this genus of flower, which is experimentally verifiable, one can then do the inverse, namely the controllability issue of creating desired flower patterns. That also is the essence in cancer dynamics, namely of creating a control or cure, but with a verifiable model. One must have the model, thus say the authors. Thus says nature! If one takes the authors systems approach and applies it to intercellular systems thinking, then it can be argued that the stem cell of cancer

[20] Dougherty p 166.

[21] Dougherty p 171.

theory as has been recently evoked can be readily explained, as readily as those zebra stripes! That is the strength of the model posited by the authors.

9.4 FINAL OBSERVATIONS

Let us make a few final observations regarding this approach. We have argued before that we approach this effort as an engineer, not as a bench Biologist. Our primary interest is in condensing experimental facts into a systems model upon which we can expand powerful tools to gain insight into diagnostic and prognostic levels as well as into ways to control and arguable cure the cancers we have been examining.

There are approaches which we believe may have fatal flaws. Let us examine a few and discuss why we have concern.

1. GWAS: Genome wide expression analyses have led to a massive explosion in gene profiles that are putatively linked to various disorders from cancers to diabetes. However the strength of such relationships is weak at best and confusing at worst. It is particularly confusing to patients who ask for tests and their physicians are confused as to the usefulness of the complex results.

2. Microarray: Microarrays as we have discussed them are likewise a tool that suddenly and inexpensively allows "discovery" of genetic relationships. We have the same concern here as we do with other approaches.

3. Correlative Gene Expressions: Using some of the above tools we have seen multiple gene expressions used for prognostic tools. We have recently discussed this for prostate cancer comparing two tools. In only one gene was there a semblance of similarity. In neither case do we have an arguable causal relationship. It is nearly impossible to argue "why" and moreover to demonstrate how these genes operate.

The recent report on such a cancer prognostic model such as Oncotype DX by Knezvic et al is a putatively prognostic method used in prostate cancer. Fundamentally what they do is examine cancer cells for the expression of various genes and examine three sets; baseline expressions, excess expressions and reduced expression. They use the baseline to set levels for excess and reduced. They then use the excess or reduced in a one dimension expression to determine a prognostic measure. This seems to be in contrast with work we reported on a few months ago[22].

Like PSA measures, CA125, CEA, and the like, they try to reduce everything to a single number. We argue here that such an approach is problematic at best. Furthermore they fail totally to demonstrate any internal pathway influence. There is no predictive basis for their approach predicated upon the actual dynamics of the cell. It is purely correlative and there may be substantial confounders involved. This approach is an example of what we fell to be the poorer aspects of genomics applied to cancer prognostics.

[22] See Knezvic, D., et al, Analytical validation of the Oncotype DX prostate cancer assay – a clinical RT-CR assay optimized for prostate needle biopsies, BMC Genomics 2013, 14:690.

In a recent study the authors develop a score called the GPS score which is based upon know malignant PCs cells and then argue that then score has significant prognostic value. The authors state:

The Oncotype DX Prostate Cancer Assay has been clinically validated, demonstrating that the GPS, assessed in diagnostic biopsy tissue, can predict the likelihood of the presence of adverse pathology (high-grade and/or high-stage disease), and that it complements existing pre-treatment risk assessment tools such as PSA levels, Gleason Score, and clinical stage.

Thus the approach contained herein is clearly causal. It is structural and it reflects a truly modellable and predictable methodology.

Now there are many significant issues in this analysis.

1. The weights are arguably chosen to maximize the risk of missing an aggressive PCa. However I have not yet seen adequate clinical evidence to that effect.

2. Prior proposed genes and the ones included herein are shown below, one from the study currently in discussion and the other from a prior study of a Myriad genetic profile:

3. In the current test under discussion the cells used for extraction are arguably from the prostate biopsy. The Myriad appear to be more wide spread.

4. Are these tests worth anything? Furthermore, groups are offering tests to assess risks based upon genetic profiles.

Just because some genetic profile may have some correlative relationship the genetic profile is not causative. Tests like these can be costly and of yet to be fully justified clinical value. Take a melanoma, if one has a suspect pigmented lesion then a simple excision and competent path study should suffice. That is an order of magnitude less than the genetic profile. In fact if one were to do a profile it should be of the melanocytes and not of the cells in general.

9.5 SUMMARY

Let us make a final summary of the issues we brought out in this paper. In a NEJM article is was recently noted[23]:

This is the age of massive genome surveys — at least for a little while longer. Sixty years after Watson and Crick's discovery, and a decade after the completion of the Human Genome Project, large-scale sequencing efforts directed at human disease abound, especially for cancer and rare congenital syndromes. International research teams supported by public funding agencies such as the National Institutes of Health and by private foundations such as the Wellcome Trust are rapidly enlarging the catalogue of genetic changes associated with neoplasia and other ailments, using ever faster, ever cheaper sequencing methods and heavy-duty bioinformatics.

[23] http://www.nejm.org/doi/full/10.1056/NEJMe1303816

Critics of big genomics projects have argued that such work is resource-intensive, is not hypothesis-driven, and amounts to little more than molecular philately. But as discoveries from these projects stack up, and as terabytes of observational data yield new insights into disease biology and prompt the development of pathway-driven targeted therapies, the usefulness of such approaches is becoming undeniable. When the Cancer Genome Atlas (TCGA) wraps up its 8-year effort next year, it will have provided detailed information on 10,000 cancer genomes for less than the cost of a trio of F-22 Raptor stealth fighters.

Let us examine where and where "not" models function. The issue we are interested in is that to develop models and substantiated inferences we need a well understood reality for cause and effect. Namely gravity causes a force between two masses, it is a measured effect. Charge causes a force between two particles, it is a measured effect. In gene structures PTEN modulates PI3K, BRAF activates MEK and the pathways are well known. We know ligands, we know receptors, and we know gene activators.

We also understand how these function, the forces, charges, conformations. We know how to inhibit and activate. We know these by facts not by correlations. Thus when we know these facts we have a basis for, indeed a demand for, using these dynamic models as an integral part of our understanding. We cannot and must not resort to random correlations. At times everything can be correlated.

Let me first explain by examining what human intellect has developed using models like this and then examine those things where our knowledge is severely lacking. The driver for this is that genomics is more akin to physical systems which we can do a great deal with and NOT akin to Economics and its correlations which frankly led us nearly to an economic collapse.

Inherent in each of the areas where the use of the knowledge of relationships is integral to the descriptions, and not correlations, is also simple yet high level descriptive; Schrodinger's Equation, Navier Stokes Equation, Fokker Plank Equation, Kushner Stratonovich Equation. We have argued elsewhere that for genomic systems we are almost already there, just a few more steps.

9.5.1 Where They Work

Let us first examine where these principles work:

Thermodynamics

Early 19th century thermodynamics did not understand the true nature of heat, namely the movement of molecules and the statistical behavior. From the gross concepts we arrived at such things as internal energy, heat, enthalpy and other gross measures of a system's thermodynamic properties. With the development of statistical mechanics there was the move from gross measurement to understanding the statistical distribution of the molecules and this was presented via the Fokker Planck equation, a means to examine the detailed statistical fine structure of complex mixtures with thermodynamic issues.

Fluid Flow

Understanding fluid dynamics was initially a study with many tables and data taken from past experiments. Slowly as the theory evolved the Navier Stokes theory came forth and constructs such as flow fields evolved and then random changes in turbulence theory also was developed.

Stochastic Dynamic Systems

Complex systems, namely entities which are based on physical realities, aircraft control systems, chemical process controls, and the like can be modeled by a complex multidimensional spatio temporal model. Applying statistical methods developed by Wiener and Kolmogoroff one saw the development of the Kushner Stratonovich equations and then saw them extended to distributed system as well. This analysis allows one to analyze highly complex multidimensional stochastic systems in time and space.

Wiener considered these in his studies of Cybernetics, and furthermore it was Wiener who also started the development of understanding complex organic systems.

The most important elements in using system models is our ability to ascertain Observability and Controllability; the ability to reproduce the model and the ability to send the states in the model to a desired end point. We also need to have the ability to identify the coefficients in the model. We often have good initial guesses but having measurement means that we can continually refine them

Electronics

When the transistor was invented the manager of the people who did it promptly published a book on solid state theory. Very few had a clue as to what Shockley was saying and frankly for those who used the device no one really cared. The electronics designer knew the linkages, and how to modify the, A good electronics designer knew that if this voltage went up the other went down, or whatever.

One knew the equations of the voltages and currents, one understood the complexities of the time varying circuits. There was a substantial amount of well proved physics behind all of this. However a good engineer also understood the ebb and flow, as for example a good neurologist can be examining the patient understand where the lesion is and then find out what the lesion is.

We are starting to get to that point with genomics. We know if we activate a kinase receptor then we activate a certain set of pathways which activate a certain set of transcription factors. A good Genomics Engineer does not need to “know” the protein structures, just what they do, at a very high level, yet detailed enough to catch the unique events which may occur.

Quantum Mechanics and Electrodynamics

Erwin Schroedinger came up with a simple manifestation of electrons in a quantum world. Feynman came up with simple diagrams to show what sub atomic particles are doing when they

interact. Now solving the Schroedinger equation for a complex organic molecule is not readily achieved it can be attacked using sophisticated computers.

9.5.2 Where The Models do NOT Work:

Now let us examine where they do not work:

Economics

Economics pretends to be mathematical. In reality, other than the tautologies of balancing financial sheets, the demand and supply theories are pure abject speculation. Econometrics is merely a fallacious use of correlation theory. There is no fundamental cause and effect, no demonstrable demonstration between input and output. This should be a warning for those working in Genomics. Just having a correlation is not causality and furthermore there is an underlying reality that is being ignored.

Social Sciences

The social sciences have tried to ascertain human responses. Approaches such as those used in election prediction may function in the short term but humans are all too often less than a herd and change opinions all too frequently.

Psychology

Psychology is strewn with the dead bodies of mathematical approaches to understanding human behavior. The problem is that we do not understand the brain, yet, and thus models of thinking, such as those in artificial intelligence are at best primitive guesses.

9.5.3 Fundamental Requirements

To have a Genomic Model or something useful we must have the following:

1. A Verifiable Realization of How the System Works. This entails the understanding of pathways and their effects on cell proliferation and movement.

2. Some Understanding of What Causes Changes in Pathways. Here we have a difficulty. We not only have somatic changes, but we have epigenetic changes such as micro RNAs and even methylation and the like. We also must understand germ line predispositions. These are gene predispositions inherited as well as SNP predispositions which can result in subsequent translation of proteins.

3. Environmental Understandings: This would include the extracellular matrix issues and its environment as well as the ability of the invading malignant cells to activate surrounding benign cells to assist in proliferation.

4. An Acceptable Measure of the Malignant Cell in Space and Time. As with the Fokker Planck equation or the Kushner Stratonovich model this may mean a measure such as the average number of a specific type of malignant cell per volume at each spatio-temporal location.

5. A Realization for the Progression of Somatic Changes: This means an understanding of the statistical nature of the somatic genetic changes as cells progress in time. For example what happens when we go from MDS to AML. This is not AML de novo. We do not need the details but the transition rates and the possible states. This issue is akin to the electronics world where we need higher level understanding and not the details.

6. We Need Ability to Integrate Parameter Identification with Stochastic Models: Clearly if we know the models and if we know that certain factors are the drivers of these models, we may use this to identify the parameters at the same time we are estimating the states.

7. We need a Methodology to Quantify Our Representations and to validate them: This is akin to having a Kushner Stratonovich equation. It is what we have developed by using average number of cells by type at specific spatio-temporal sites. I believe we have the techniques, they are built on the many other approaches.

9.5.4 The Genomic Model

The Genomic Model is a systematizeable model. It is s system wherein we have well known causes and effects. We know that if a ligan attaches to a receptor then one has an activated pathway that induces a transcription factor which results in cellular proliferation. We know cause and effect. We know the rates of these factors. We can also develop models wherein we can estimate the average number of cells of a particular genetic structure at a specific time and at a specific place in the body.

The NEJM article concludes:

In 1803, a few years before the inaugural issue of the Journal, Thomas Jefferson commissioned Meriwether Lewis and William Clark to survey the vast unknown American frontier. Lewis and Clark departed from St. Louis, where Ley et al. initiated the AML genome survey. Less than a century later, the western frontier was declared "closed," but land surveyors did not disappear; today, they focus on construction projects and property boundaries. Likewise, although the initial epic AML genomic survey that began in St. Louis is now largely complete and surveys of other neoplasms will soon conclude, the use of genomics in quotidian practice is just beginning.

Now if we were to interpret this correctly it sends just the wrong message. The developments in Genomics are not Lewis and Clark like, they require Newtonian and Maxwellian insight. Fundamental laws and relationships, causality, albeit stochastic, and determine ng the right measures.

Thus in a sense one could imagine the Genomics Engineer being akin to the electronics engineer, or even to the neurologist. As many a medical student would recall from anatomy, the tracing of the cranial nerves is a critical skill, but one enhanced by seeing how they migrate from back to

front. Diagnosis of cranial nerve issues are resolved by understanding the network. In a similar manner we would hope the same is double with Genomics pathways,

10 REFERENCES:

1. Ahn, K., Transcription Factor NF-kB, Ann NY Aca Sci, V 1056, 2005, pp 218-233.
2. Ailles, L., I. Weissman, Cancer Stem Cells in Solid Tumors, Curr Opp in Biotech, 2007, pp 460-466.
3. Allis, C., et al, Epigenetics, CSH (Cold Spring Harbor) 2007.
4. Ando, S., H. Iba, Quantitative Modeling of Gene Regulatory Network, Univ Tokyo.
5. Appasani, K., MicroRNAs, Cambridge (New York) 2008.
6. Araud, T., Overview of the miRNA Pathways, Thesis, U Geneva, 2008.
7. Athans, M. et al., System, Networks, & Computation; Multivariable Methods, McGraw Hill (New York) 1974
8. Baici, A., *Enzyme Kinetics*, The Velocity of Reactions, Biochem Journal, 2006, pp. 1-3.
9. Baker, C.A.H., et al, GeneVis: Visualization Tools for Genetic Regulatory Network Dynamics, Univ Calgary.
10. Baldo, P., et al, mTOR Pathway and mTOR inhibitors as Agents for Cancer Therapy, 2008, Curr Can Drug Targets, pp 647-668.
11. Bartel, D., MicroRNAs: Genomics, Biogenesis, Mechanism, and Function, Cell 2004 pp 281-297.
12. Bellomo, N., et al, Selected Topics in Cancer Modeling, Birkhauser (Boston) 2008.
13. Benson D. L., Unraveling the Turing Bifurcation Using Spatially Varying Diffusion Coefficients, Jour Math Bio, Vol 37 1998, pp. 381-417.
14. Bolouri, H., Computational Modeling of Gene Regulatory Networks, Imp Coll Lon Press (London) 2008.
15. Bunz, F., Principles of Cancer Genetics, Springer(New York) 2008.
16. Calin, G., C. Croce, Chromosomal Rearrangements and Micro RNAs, Jrl Clin Inv 2007 pp 2059-2066.
17. Campbell, A., L. Heyer, **Genomics, Proteomics, and Bioinformatics**, Benjamin Cummings (New York) 2003.
18. Carrecedo, A., et al, PTEN Level in Tumor Suppression, How Much is Too Little?, *Cancer Res; 71(3); pp. 629–33. 2011.*
19. Causton, H. et al, **Microarray Gene Expression and Analysis**, Blackwell (Malden, MA) 2003.
20. Chen, P., J. Chen, A Markovian Approach to the Control of Genetic Regulatory Networks, Bio Sys, 2007 PP 535-545.

21. Chen, T., et al, Modeling Gene Expression with Differential Equations, Pacific Symposium on Biocomputing, 1999 pp. 29-40.
22. Chen, T., et al, Modeling Gene Expression with Differential Equations, Pacific Symposium on Biocomputing, 1999 pp. 29-40.
23. Chow, L. et al, PTEN Function in Normal and Neoplastic Growth, Cancer Ltr, V 241, 2006, pp 184-196.
24. Ciechanover , A, Intracellular Protein Degradation: From a Vague Idea through the Lysosome and the Ubiquitin-Proteasome System and onto Human Diseases and Drug Targeting, Rambam Maimonides Medical Journal, January 2012, Volume 3, Issue 1.
25. Costa, F., et al, Concise Review: Cancer Testis Antigens, Stem Cells and Cancer, Stem Cells 2007 pp 707-711.
26. Crampin, E., et al, Extracting Biochemical Reaction Kinetics from Time Series Data, LNAI 3214 pp 329-336 2004 Springer Verlag.
27. Croce, C., Causes and Consequences of microRNA Dysregulation in Cancer, Nature 2009 pp 704-714.
28. Dahia, P., PTEN, A Unique Tumor Suppressor Gene, Endo Rel Can, V 7 2000, pp 115-129.
29. Dahlman, A., et al, Effect of Androgen Deprivation Therapy on the Expression of Prostate Cancer Biomarkers, Pros Can and Prost Dis 2010 pp 1-7.
30. Dalerba, P., et al, Cancer Stem Cells: Models and Concepts, Stem Cells, 2008 pp 267-284.
31. DeVita, V., et al, Cancer, Lippincott (Philadelphia) 2008.
32. di Bernardo, D., et al, Chemogeneomic Profiling on a Genome Wide Scale Using Reverse Engineered Gene Networks, Nature Biotech, V 23, 2005, pp 377-383.
33. Djebbari, A., J. Quackenbush, Seeded Bayesian Networks, BioMed Cent Sys Bio 2008 pp 1-13.
34. Dougherty, E., Epistemology and the Role of Mathematics in Translational Science, To Be Published, Texas A&M.
35. Dougherty, E., et al, Epistemology of Computational Biology, Jour Bio Sys, V 14, 2006, pp 65-90.
36. Dougherty, E., et al, Validation of Computational Methods in Genomics, Curr Gen, V 8, 2007, pp 1-19.
37. Dougherty, E., On The Epistemological Crisis in Genomics, Current Genomics, V 9, 2008, pp 69-79.
38. Dougherty, E., Translational Science: Epistemology and the Investigative Process, Curr Gen, V 10, 2009, pp 102-109.
39. Driscoll, M., T. Gardner, Identification and Control of Gene Networks in Living Organisms, Jrl Proc Contrl 2006 pp 303-311.
40. Duda, R., et al, Pattern Classification, Wiley (New York) 2001.
41. Easton, J., P. Houghton, mTOR and Cancer Therapy, Oncogene, 2006, pp 6436-6446.

42. Esteller, M., Epigenetics in Cancer, NEJM, 2008 V 358 pp 1148-1159.
43. Feero, W. G., et al, Genomic Medicine – An Updated Primer, NEJM, 2010, pp 2001-2011.
44. Finn, O., Cancer Immunology, NEJM V 358, 2008 pp 2704-2715.
45. Frohling, S., Chromosomal Abnormalities in Cancer, NEJM 2008, pp 722-734.
46. Gardenes, J., et al, On the robustness of complex heterogeneous gene expression networks, Biophys Chem, 118 2005 pp. 225-228.
47. Gershenson, C., Classification of Random Boolean Networks, in Artificial Life, MIT Press 2002., pp 1-8.
48. Glunde, K. et al. Molecular–functional imaging of cancer: to image and imagine, TRENDS in Molecular Medicine, Vol.13 No.7, 2007.
49. Goentoro, L., et al, The Incoherent Feedforward Loop Can Provide Fold Change Detection in Gene Regulation, Mol Cell 2009 pp 894-899.
50. Goldberg, A., Protein degradation and protection against misfolded or damaged proteins, NATURE, Vol 426, 18/25 December 2003.
51. Gottwein, E., B. Cullen, Viral and Cellular MicroRNAs as Determinants of Viral Pathogenesis and Immunity, Cell Host 2008, pp 375-387.
52. Griffiths, A., et al, **Genetic Analysis 5th Ed**, Freeman (New York) 1993.
53. Grochow, J., M.. Kellis, Network Motif Discovery Using Subgraph Enumeration and Symmetry Breaking, Speed and Huang Ed, RECOMB Springer 2007 pp 92-106.
54. Hanley, J., B., McNeil, The Meaning and Use of the Area under a Receiver Operating Curve, Radiology 1982 V 143, pp 29-36.
55. Hatzimanikatis, V., Dynamical Analysis of Gene Networks Requires Both mRNA and Protein Expression Information, Metabolic Engr, Vol 1, 1999, pp. 275-281.
56. Hay, N., The Akt-mTOR tango and its relevance to Cancer, Cancer Cell, V 8, 2005, pp 179-183.
57. He, L., G. Hannon, MicroRNAs: Small RNAs with a Big Role in Gene Regulation, Nature, 2004 pp 522-532.
58. Herman, J., S. Baylin, Gene Silencing in Cancer in Association with Promoter Hypermethylation, NEJM 2003, pp 2042-2034.
59. Holmes, S., Graphs and Networks in Bioinformatics, Stanford Univ Notes.
60. Hosmer, D., S., Lemeshow, Applied Logistic Regression, Wiley (New York) 2000.
61. Houston, P., Chemical Kinetics and Reaction Dynamics, Dover (Mineola) 2001.
62. Hsu. P, David M. Sabatini1,2,3 Cancer Cell Metabolism: Warburg and Beyond, Cell 134, September 5, 2008.
63. Huang, C., et al, A Study of Cancer Related MicroRNAs through Expression Data and Literature Search, World Acad Sci & Engr, 2009 254-256.
64. Huang, S., et al, Blockade of NF-kB Activity in Human Prostate Cancer Cells is Associated with Suppression of Angiogenesis, Invasion, and Metastasis, Oncogene, V 20, 2001, pp 4118-4197.

1. Huettel, S., et al, Functional Magnetic Resonance Imaging, Sinauer (Sunderland, MA) 2009.

65. Hunink, M., C. Begg, Diamond's Correction Method, A Real Gem or Just a Cubic Zirconium, Med Dec Making 1991 V 11 pp 201-203.

66. Imoto, S., et al, Combining Microarrays and Biological Knowledge for Estimating Gene Networks via Bayesian Networks, IEEE Com Sys Bio 2000 pp .

67. Ivanov, I., E. Dougherty, Reduction Mappings Between Probabilistic Boolean Networks, Jrl App Sig Pro 2004 pp 125-131.

68. Iwarta, T., et al, Myc Overexpression Induces Prostatic Intraepithelial Neoplasia and Loss of Nkx3.1 in Mouse Luminal Epithelial Cells, Open Access, www.plosone.org ,V 5, 2010, pp e9427

69. Jacobsen, S., Gene Silencing: Maintaining Methylation Patterns, Curr Bio 1999, pp 617-619.

70. Jansen, R., et al, A Bayesian Networks Approach for Predicting Protein-Protein Interactions from Genomic Data, Science, 2003 pp 449-453.

71. Jones, P., S. Baylin, The Epigenomics of Cancer, Cell, 2007 pp 683-692.

72. Jordan, C., et al, Cancer Stem Cells, NEJM 2006, pp 1253-1262.

73. Kanamori, Y., et al, Correlation Between Loss of PTEN Expression and Akt Phosphorylation in Endometrial Carcinoma, Clin Can Res, V 7, 2001, pp 892-895.

74. Kilkov, V., Identifying Gene Regulatory Networks from Gene Expression Data, .

75. Kim, S., et al Can Markov Chain Models Mimic Biological Regulation, Jrl Bio Sys 2002, pp 337-357.

76. Kim, S., et al, Inferring Gene Networks from Time Series Microarray Data Using Dynamic Bayesian Networks, Brief on Bioinfo 2003, pp 228-235.

77. Klemm, K., S. Bornholdt, Topology of Biological Networks and Reliability of Information Processing, PNAS 2005 pp 18414-18419.

78. Klipp, E., et al, Systems Biology, Wiley (Weinheim, Germany) 2009.

79. Knezvic, D., et al, Analytical validation of the Oncotype DX prostate cancer assay – a clinical RT-CR assay optimized for prostate needle biopsies, BMC Genomics 2013, 14:690.

80. Kohane, I., et al, **Microarrays for an Integrative Genomics**, MIT Press (Cambridge) 2003.

81. Krane, D., M. Raymer, Bioinformatics, Benjamin Cummings (New York) 2003.

82. Krause, D., R. VanEtten, Tyrosine Kinases as Targets for Cancer Therapy, NEJM 2005 pp 172-187.

83. Lam, E., et al, FOXO Transcription Factors, Biochem Soc Trans 2006, pp 722-726.

84. Lee, A., GRP78 Induction in Cancer, Can Res 2007 pp 3946-3499.

85. Lesk, A., Bioinformatics, Oxford (New York) 2002.

86. Levine, A., A. Puzio-Kuter, The Control of the Metabolic Switch in Cancers by Oncogenes and Tumor Suppressor Genes, Science, 2010, pp 1340-1344.

87. Li, G., et al, Selective Anticancer Strategies via Intervention of the Death Pathways Relevant to Cell Transformation, Cell Death and Diff 2008 pp 1197-1210.

88. Ljung, L. System Identification, Prentice Hall (Englewood Cliffs) 1987.

89. Lobo, N., et al, The Biology of Cancer Stem Cells, Ann Rev Cell Dev 2007 pp 675-699.

90. Lodish, H., et al, Molecular Cell Biology, 4th Ed Freeman (San Francisco) 2000.

91. LoPiccolo, J., et al, Targeting the PI3K/Akt/mTOR Pathway, Drug Res Up, 2008, pp 32-50.

92. Lu, Y., et al, A Bayesian Approach to Simultaneously Adjusting for Verification, Statst Med 2010 V 29, pp 2532-2543.

93. Maira, S., et al, PI3K Inhibitors for Cancer Treatment, Biochem Soc Trans, V 37, 2009, pp 265-272.

94. Marks, F., et al, Cellular Signal Processing, Garland (New York), 2009

95. Martinez-Vincente, M., et al, Protein degradation and aging, Experimental Gerontology 40 (2005) 622–633.

96. McGarty, T. Achieving the "Blue Daylily": Genetic Capabilities for Color in the Genus Hemerocallis, Draft, MIT, August 2010

97. McGarty, T. Backcrossing: A Mathematical Analysis of Gene Insertion in Existing Hybrids and a Statistical Validation, Draft, MIT August 2008

98. McGarty, T. P, The Structure of the Upper Atmosphere, IEEE Automatic Control 1971.

99. McGarty, T. P., Bayesian Outlier Rejection and State Estimation, IEEE AC 1975, pp. 682-687

100. McGarty, T. PSA Evaluation Methodologies: A Look at Multiple Alternatives and Maximum Likelihood Techniques, Draft, MIT, December 2010.

101. McGarty, T. The PSA Controversy: Details, Models, Analysis and Recommendations, Draft, MIT, December 2010.

102. McGarty, T., Flower Color and Patterning in the Genus Hemerocallis and its Hybrids: A Mathematical Model and Experimental Analysis , MIT, Draft, August 2009.

103. McGarty, T., Health Care Policy, Telmarc Press (New York) 2009.

104. McGarty, T., Gene Expression in Plants: Use of System Identification for Control of Color, submitted to IEEE JSAC for 2008.

2. McGarty, T., Melanoma System Genomics, DRAFT, 2013.

3. McGarty, T., Prostate Cancer Genomics, DRAFT, 2013.

4. McGarty, T., Stochastic Systems and State Estimation, Wiley (New York) 1974.

105. McGarty, T.P., The PSA Controversy: Details, Models, Analysis and Recommendations, November 2010, MIT Working Paper.

106. McIntosh, M, N., Urban, A Parametric Empirical Bayes Method for Cancer Screening, Biostat 2003 V 4, pp 27-40.

107. McMenamin, M., et al, Loss of PTEN Expression in Paraffin-embedded Primary Prostate Cancer, Can Res, V 59, 1999, pp 4291-4296.

108. McMurry, J., Begley, T., The Organic Chemistry of Biological Pathways, Roberts & Company Publishers, 2005.

109. McNeil, B., J., Hanley, Statistical Approaches to the Analysis of Receiver Operating Curves, Med Dec Making V 4 1984 pp137-149.

110. Mello, C., Return of the RNAi World: Rethinking Gene Expression and Evolution, Nobel Laureate Lecture 2006.

111. Metzker, M., Sequencing Technologies, Nature Rev V 11, 2010, pp 31-46.

112. Milo, R., et al, Network Motifs: Simple Building Blocks of Complex Networks, Science, 2002, pp 824-827.

113. Miranda, T., P. Jones. DNA Methylation: The Nuts and Bolts of Repression, Cll Phys 2007 pp 384-390.

114. Mjolsness, E., et al Gene Regulation Networks with Signaling and Cell Division in Development Satiations, IEEE Biomed Engr Houston TX Oct 2002.

115. Mooi, W., D. Peeper, Oncogene Induced Cell Senescence, NEJM, 2006 pp 1037-1045.

116. Morgan, D., The Cell Cycle, Sinauer (London) 2007.

117. Mosquera, J., et al, Morphological Features of TMPRSS2-ERG Gene Fusion Prostate Cancer, Jrl Path 2007 pp 91-101.

118. Mosquera, J., et al, Prevalence of TMPRSS2-ERG Fusion Prostate Cancer among Men Undergoing Prostate Biopsy in the United States, Clin Can Res, V 15, 2009, 4706-4711.

119. Murphy, K., S. Mian, Modelling Gene Expression Data Using Dynamic Bayesian Networks, .

120. Murray, J., **Mathematical Biology**, Springer (New York) 1989.

121. Murrell, J., Understanding Rate of Chemical Reactions, University of Sussex.

5. Murtaza M et al, Noninvasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA (2013) Nature.

122. Nelson, E., et al, Inhibition of Akt Pathways in the Treatment of Prostate Cancer, 2007, Pros Can and Pros Dis, pp 331-339.

123. Nusse, R., Relays at the Membrane, Nature, 2005, V 438, pp 747-748.

6. Ozkumur, E., Inertial Focusing for Tumor Antigen–Dependent and –Independent Sorting of Rare Circulating Tumor Cells, Sci Transl Med 3 April 2013: Vol. 5, Issue 179, p. 179.

124. Pali, S., K. Robertson, Epigenetic Control of Tumor Suppression, Clin Rev in Euk Gene Exp 2007, pp 295-316.

125. Papatsoris, A., et al, The Power and Promise of "Rewiring" the Mitogen Activated Protein Kinase Network in Prostate Cancer Theraputics, Mol Can Ther, V 6, 2007, pp 811-819.

126. Payne, S., C. Kemp, Tumor Suppressor Genetics, Carcinogenesis, V 26, 2005, pp 2031-2045.

127. Pecorino, L., Molecular Biology of Cancer, Oxford (New York) 2008.

128. Peng, C., et al, An Introduction to Logistic Regression Analysis and Reporting, Jrl Ed Res, V 96 2002 pp 3-9.

129. Pepe, M., The Statistical Evaluation of Medical Tests for Classification and Prediction, Oxford (New York) 2003.

130. Percus, J., Mathematics of Genome Analysis, Cambridge (New York) 2004.

131. Perkins, T., et al, Inferring Models of Gene Expression Dynamics, Journal of Theoretical Biology, Vol 230, 2004, pp. 289-299.

132. Planet, P. et al, Systematic Analysis of DNA Microarray Data: Ordering and Interpreting Patterns of Gene Expression, Cold Spring Harbor, Genome Research, 2001 pp 1149-1155.

133. Poliseno, L., et al, Identification of the miR-106b MicroRNA Cluster as a Proto Oncogenic PTEN Targeting Intron That Cooperates with the Host Gene MCM7 in Transformation, Sci Sig 2010, pp 29- .

134. Punglia, R., et al, Effect of Verification Bias on Screening for Prostate Cancer, NEJM, 2003 V 349, pp 335-342.

135. Putzi, M., A. De Marzo, Morphologic Transitions Between Proliferative Inflammatory Atrophy and High Grade Prostatic Intraepithelial Neoplasia, Adult Uro 2000 pp 828-832.

136. Radu, A., et al, PTEN Induces Cell Cycle Arrest by Decreasing the Level and Nuclear Localization of Cyclin D1, Mol and Cell Bio, V 23, 203, pp 6139-6149.

137. Reya, T., et al, Stem Cells, Cancer, and Cancer Stem Cells, Nature, 2001 pp 105-111.

138. Robertson, K., A. Wolffe, DNA Methylation in Health and Disease, Nat Rev Gen 2000 pp 11-19.

139. Robertson, K., DNA Methylation, methyltransferase and cancer, Oncogene 2001 pp 3139-3155.

140. Rosen, J., C. Jordan, The Increasing Complexity of the Cancer Stem Cell Paradigm, Science, 2009, pp 1670-1672.

141. Rosenberg, S., Gene Therapy for Cancer, JAMA 1992 pp 2416-2419.

142. Rosenberg, S., The Transformed Cell, Putnam (New York) 1992.

143. Roy, S., et al, Inhibition of PI3K/AKT and MAPK/ERK pathways causes Activation of FOXO Transcription Factor, Jrl Mol Sig 2010, pp 1-13.

144. Rubin, M., A. Chinnaiyan, Bioinformatics Approach Leads to the Discovery of the TMPRSS2:ETS gene Fusion in Prostate Cancer, Lab Inv 2006, pp. 1099-1102.

145. Rubinow, S., Mathematical Biology, Dover (New York) 2002.

146. Sachs, K., et al, Causal Protein Signalling Networks Derived from Multiparamemeter Single Cell Data, Science, 2005, pp 523-529.

147. Sage A. P., J. Melsa, System Identification, Academic Press (New York) 1971.

148. Saitou, N., M. Nei, *The Neighbor Joining Method: A New Method for Reconstructing Phylogenetic Trees,* Molecular Biological Evolution, Vol 4 1987 pp 406-425.

149. Schnell. S, T. Turner, Reaction Kinetics in Intracellular Environments with Macromolecular Crowding, Biophys and Molec Bio vol 85 2004 pp. 235-260.

150. Segel, L., M. Slemrod, The Quasi Steady State Assumption: A Case Study in Perturbation, SIAM, 1989, pp 446-477.

151. Seoane, J., et al, Integration of Smad and Forkhead Pathways in the Control of Neuroepithelial and Glioblastoma Cell Proliferation, Cell 2004, pp. 211-223.

152. Setlur, S., et al, Integrative Microarray Analysis of Pathways Dysregulated in Metastatic Prostate Cancer, Cancer Res, V 67, 2007, pp 10296-10303.

153. Sevigani, C., et al, Mammalian microRNAs: A Small World for Fine-tuning Gene Expression, Mam Gene V 17 2006, pp 189-201.

154. Sevli, S., et al, The Function of microRNAs, small but potent molecules, in Human Prostate Cancer, Pros Can and Pros Dis, V 13, 2010, pp 208-217.

155. Sharifi, N., et al, Androgen Deprivation Therapy for Prostate Cancer, JAMA, V 294, 2005, pp 238-244.

156. Shmulevich, I, et al, Gene Perturbation and Intervention in Probabilistic Boolean Networks, Bioinfo 2002 pp 1319-1331.

157. Shmulevich, I. et al, Control of Stationary Behavior in Probabilistic Boolean Networks, J Bio Sci 2002, pp 431-445.

158. Shmulevich, I., E., Dougherty, Genomic Signal Processing, Princeton (Princeton) 2007.

159. Shmulevich, I., et al, From Boolean to Probabilistic Networks as Models of Genetic Regulatory Networks, Proc IEEE 2002 pp 1778-1792.

160. Shyam, E. et al, The ERG Gene: A Human Gene Related to the ETS Oncogene, PNAS 1987, pp 6131-6135.

161. Sica, G., Bias in Research Studies, Radiology, 2006 V 238 pp 780-789.

162. Simon, W., Mathematical Techniques for Biology and Medicine, Dover (New York) 1986.

163. Slingerland, J., M. Pagano, Regulation of Cdk Inhibitor p27 and Its Deregulation in Cancer, Jrl Cell Phys 2000 pp 10-17.

164. Spirtes, P., et al, Constructing Bayesian Network Models of Gene Expression Networks from Microarray Data, .

165. Stein, G., et al, The Molecular Basis of Cell Cycle and Growth Control, Wiley (New York) 1999.

166. Strathdee, G., R., Brown, Aberrant DNA Methylation in Cancer; Potential Clinical Interventions, Exp Rev Mol Med, 2002, pp 1-17.

167. Studier, J., K. Kappler, *A Note on the Neighbor Joining Algorithm*, Molecular Biological Evolution, Vol 5 1988 pp 729-731.

168. Styerberg, E., et al, Prediction of Indolent Prostate Cancer, Jrl Uro, 2007, V 177, pp 107-112.

169. Su, L., Early Diagnosis and Treatment of Cancer: Prostate Cancer, Saunders(New York) 2010.

170. Swanson, K, et al, Prostate Specific Antigen, Am Jrl Clin Path, 2006, V 125 pp 331-333.

171. Szallasi, Z. System Modeling in Cellular Biology: From Concepts to Nuts and Bolts. MIT Press (Cambridge) 2006.

172. Taby, R., J-P Issa, Cancer Epigenetics, CA 2010 pp376-392.

173. Tang, D., et al, Prostate Cancer Stem/Progenitor Cells, Mol Car 2007 pp. 1-14.

174. Taylor, B., Integrative Genomic Profiling of Human Prostate Cancer, Cancer Cell, V 18, 2010, pp 1-12.

175. Theodoridis, S., K., Koutroumbas, Pattern Recognition, Academic (New York) 2009.

176. Turing, A., *The Chemical Basis of Morphogenesis*, Phil Trans Royal Soc London B337 pp 37-72, 19459.

177. van der Heide, L., et al, The Ins and Outs of FOXO Shuttling, Biochem Jrl 2004, pp 297-309.

178. Vander Heiden, M., et al, Understanding the Warburg Effect, Science 2009 pp 1029-1033.

179. Vander Heiden, M., et al, Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation, SCIENCE VOL 324 22 MAY 2009.

180. Vazquez, F., et al, The PTEN Tumor Suppressor Protein, Biochimica et Biophysica, 2000 pp 21-25.

181. Vicini FA, Abner A, Baglan KL, Kestin LL, Martinez AA. Defining a dose-response relationship with radiotherapy for prostate cancer: is more really better? *Int J Radiat Oncol Biol Phys*. 2001 Dec 1;51(5):1200-8.

182. Vis, A., F. Schroder, Key Targets of hormonal treatment of prostate cancer; Part 2 the androgen receptor and 5α reductase, BJUI 2009 pp 1191-1197.

183. Vliestra, R., et al, Frequent Inactivation of PTEN in Prostate Cancer Cell Lines and Xenografts, Can Res, V 58, 1998, pp 2720-2723.

184. Vogelstein, B., K., Kinzler, Cancer Genes and the Pathways They Control, Nature Med, V 10, 2004, pp 789-799.

185. Vohradsky, J., Neural Network Model of Gene Expression, FASEB Journal, Vol 15, March 2001, pp. 846-854.

186. Wang, E., Cancer Systems Biology, CRC (New York) 2010.

187. Wang, X, et al, Notch-1 Expressing Cells are Indispensable for Prostatic Branching Morphogenesis, Jrl Bio Chem, V 279, 2004, pp 24733-24744.

188. Warburg, O., On the Origin of Cancer Cells, Science 24 February 1956, Volume 123, Number 3191.

189. Watson, J., et al, **Molecular Biology of the Gene**, Benjamin Cummings (San Francisco) 2004.

190. Weinberg, R., The Biology of Cancer, Garland (New York) 2007.

191. Werhli, A., D. Husmeier, Reconstructing Gene Regulatory Networks with Bayesian Networks, 2007 Bioinfo and Stat pp 1-45.

192. Wolf, D., et al, Transcriptional Down Regulation of c-myc in Human Prostate Carcinoma Cells by the Synthetic Mibolerone, Br Jrl Cancer, V 65, 1992, pp 376-382.

193. Wu., F.X. et al, Modeling Gene Expression From Microarray Expression Data with State Space Equations.

194. Xia, H., Great Potential of MicroRNA in Cancer Stem Cell, Jrl Can Mol 2008 pp 79-89.

195. Yang, J., et al, Activation of FOXO3 is Sufficient to Reverse Mitogen Activated Protein/Extracellular Signal, Can Res, 2010 pp 4709-4718.

196. Zak, D., Continuous Time Identification of Gene Expression Models, Journal of Integrative Biology, Vol 7 No 4 2003 pp 373-386.

197. Zhou, X., et al, Statistical Methods in Diagnostic Medicine, Wiley (New York) 2002.

198. Zilberman, D., The Evolving Functions of DNA Methylation, Curr Opin in Plt Bio, 2008, pp 554-559.