Cancer Models for Understanding, Prediction, and Control

Terrence P McGarty
White Paper No 86
June, 2012

We examine the elements required in modeling cancer and we propose several variants. We focus primarily on intracellular models using a differential equation approach although questioning and modifying the reaction rate models in common use. We also develop a model for total cellular dynamics which we integrate with the intracellular model thus allowing for whole body predictive capabilities.

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 3
2 Characteristics of Cancer 5
3 Cancer: An Evolving Puzzle 7
4 What do we Model in Cancer? 10
4.1 Intracellular Gene Dynamics: 10
4.2 Intracellular Protein Dynamics: 10
4.3 Intercellular Dynamics: 14
4.4 Total Cellular Dynamics: 15
5 Intracellular Dynamics: An Example 17
5.1 Boolean Networks 17
5.2 Bayesian Networks 20
5.2.1 Bayesian Analysis 21
5.2.2 Microarrays 24
5.2.3 Bayesian Network Summary 28
5.3 State Dynamic Networks 28
5.3.1 Kalman Filter 34
5.3.2 The Matched Filter Approach 37
6 Total Cellular Dynamics: An Example 41
6.1 Total Cellular Dynamics Models 41
6.2 Turing Tesselation 44
6.3 Determining the Coefficients 46
7 Cancer Metabolism: An Example of Complexity 48
8 Stochastic Models 51
9 Stability 52
10 Controllable and Observable 53
10.1 Controllability 53
10.2 Observability 53
11 Summary 54
12 References 55

1 CANCER MODELS FOR UNDERSTANDING, PREDICTION, AND CONTROL

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.

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 with what 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.

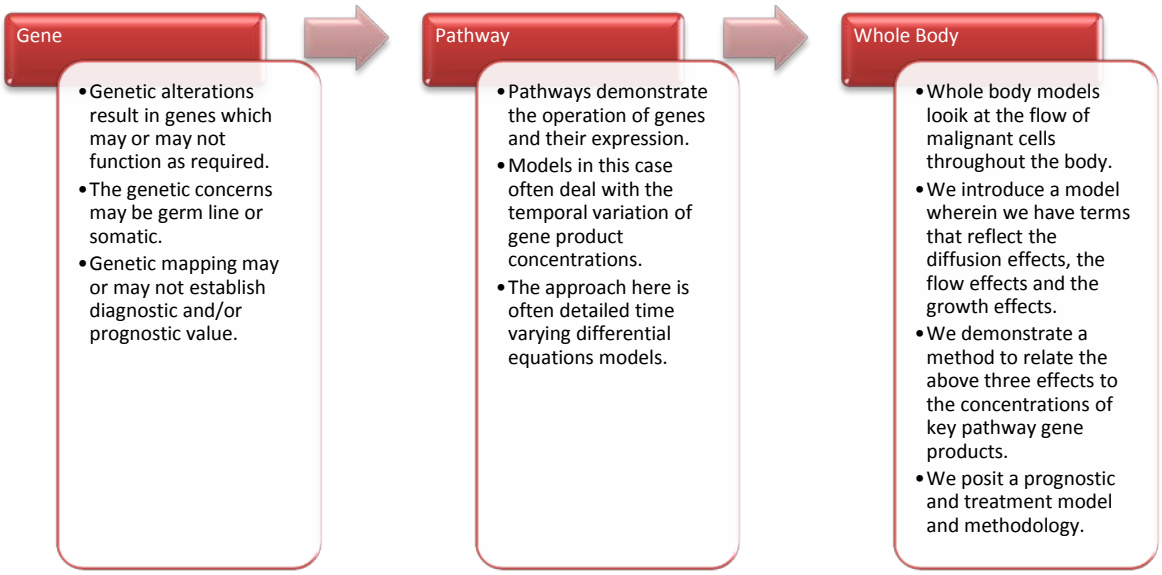
3. Measurable: One must be able to measure and then predict the quantities which make up the model. Many of the modeling systems include proteins but they react in some zero-one format. We know in reality that we have concentrations, or better yet specific numbers of proteins, produced in a cell. Yet 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.



2 CHARACTERISTICS OF CANCER

What are the key characteristics of cancer. As Weinberg notes in the figure below, normal cells have the following functions:

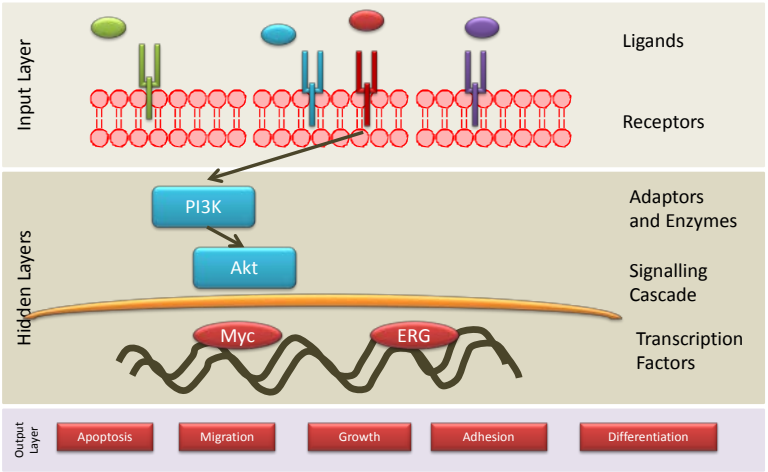
1. Growth: Growth is often distinguished from cell multiplication, or mitosis. Growth is cell expansion, the production of new proteins, the expansion of intra-cellular fluids, and the inclusion of exogenous factors found in normal cell life. In a sense cells consume and cells can get fat! The cells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

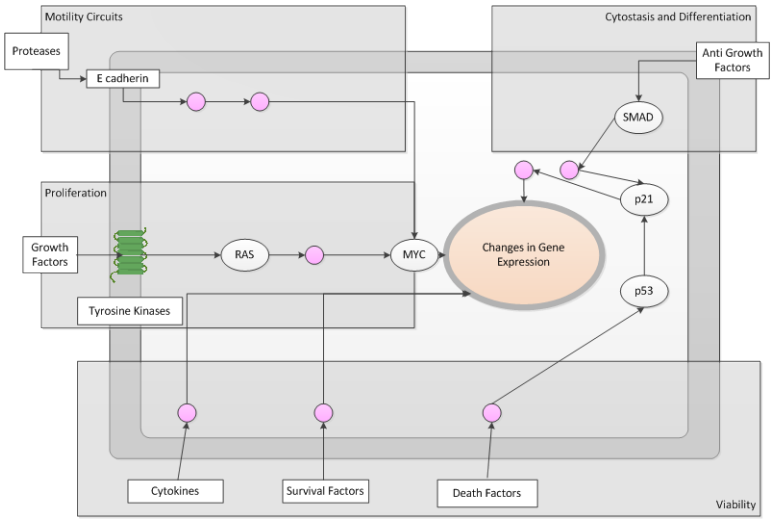
3. Loss of Normal Apoptosis: Loss of normal cell death is a major factor of cancer cells, they just do not die off.

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.

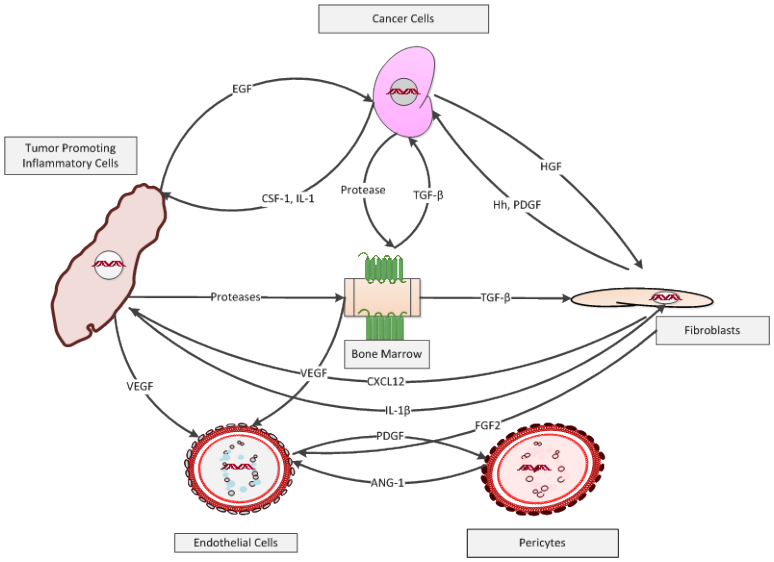
6. Development of Alternative Metabolic Capabilities: This is in effect the Warburg Effect. Namely the metabolism of the cancer cell is anaerobic rather than oxidative.

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.



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?

4 WHAT DO WE MODEL IN CANCER?

The key question we ask is just what is it we are modeling in cancer cell dynamics. Let us consider some options:

4.1 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.

4.2 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[1], 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.

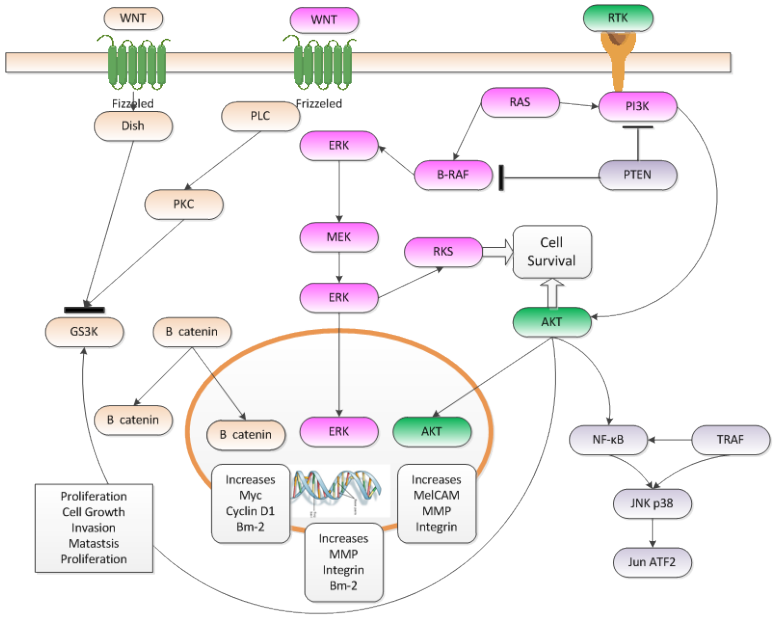
Also we refer to the recent review work of Ciechanover which details the evolution of this understanding.

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.

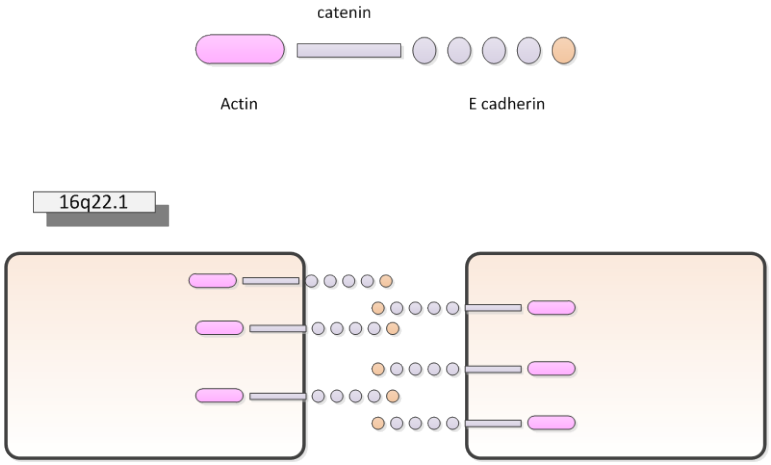
e. Total intracellular dynamics can be modeled yet the underlying processes are still not understood and the required measurements are yet to be determined.

4.3 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

4.4 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.

5 INTRACELLULAR DYNAMICS: AN EXAMPLE

We now consider in more formalistic terms how to develop models. There are three 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.

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.

5.1 BOOLEAN NETWORKS

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, 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:

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

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:

$$x_j(k+1)=(x_1(k)ANDx_3(k))NOTx_4(k)ORx_6(k)$$

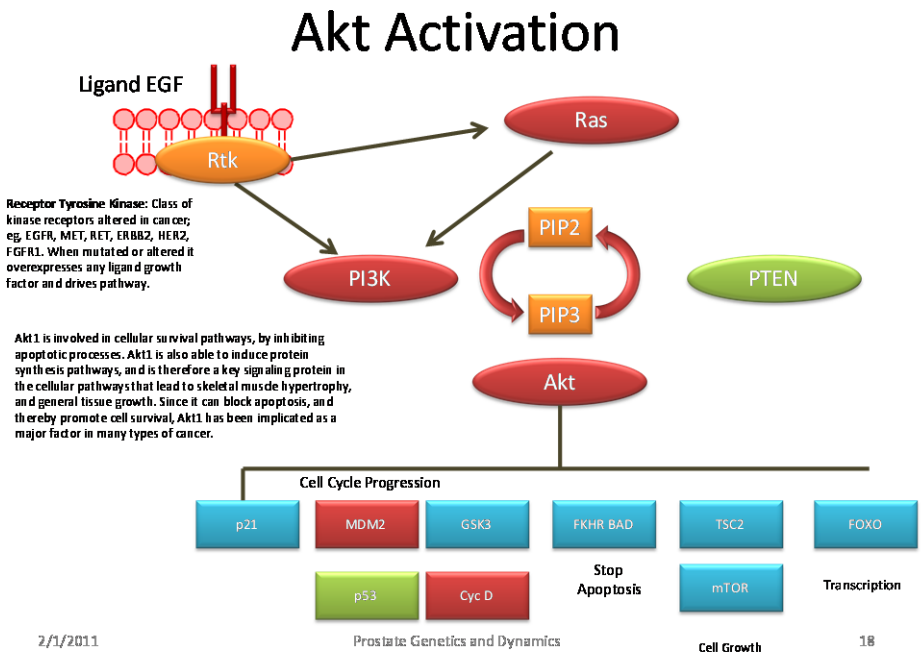
We now ask where did this equation come from. We defer that for the moment. But let us assume that we can find that for every one of the x values. The we have:

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

where

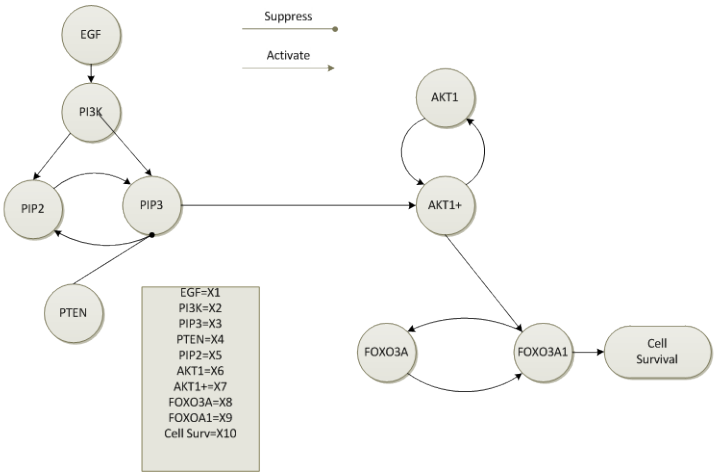
$$F(k+1,k)=\begin{bmatrix} f_1(k+1,k;x_1(k)...x_N(k)) \\ ... \\ f_N(k+1,k;x_1(k)...x_N(k)) \end{bmatrix}$$

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;



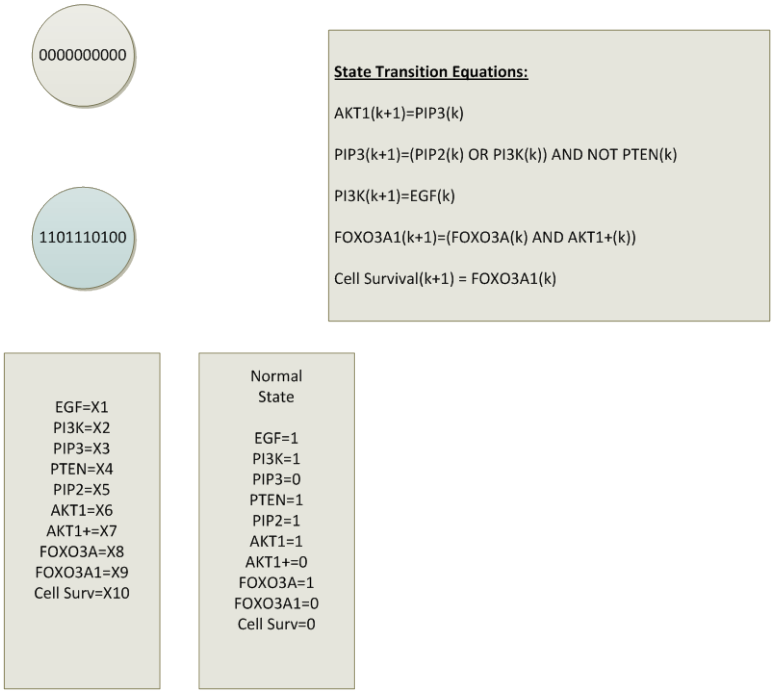
:

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)

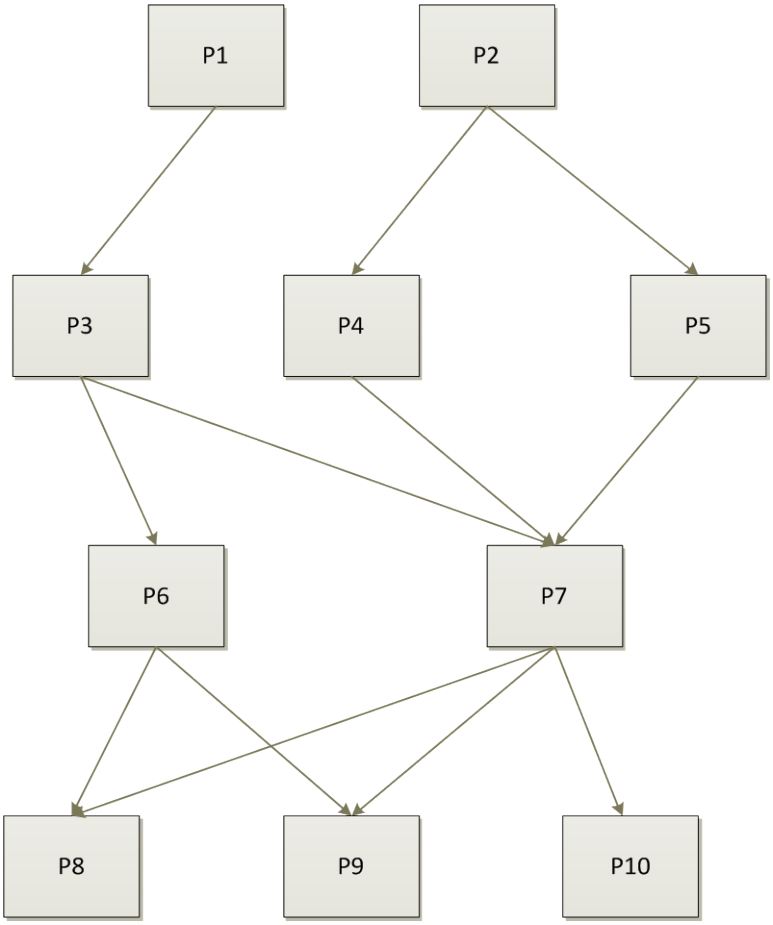
5.2 BAYESIAN NETWORKS

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.

5.2.1 Bayesian Analysis

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 was 10 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;

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

and

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

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:

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

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.

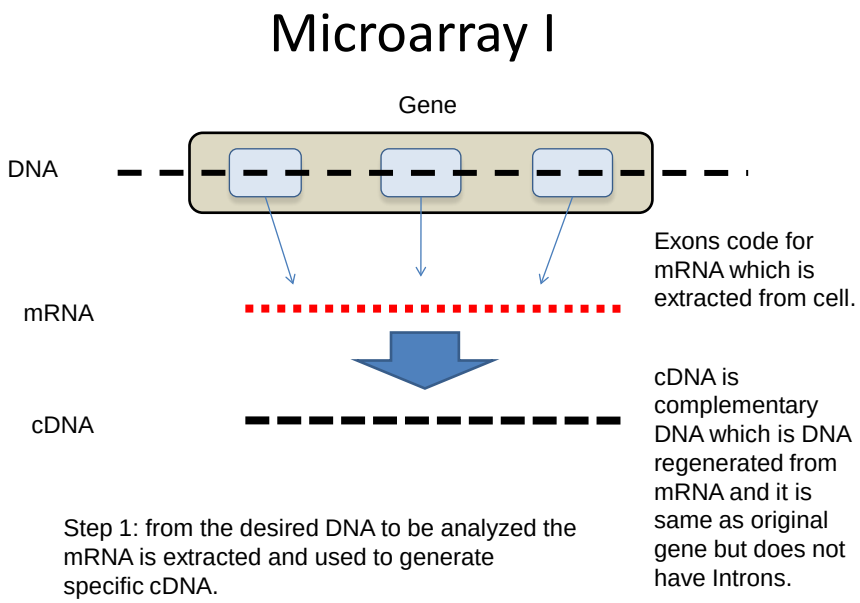
5.2.2 Microarrays

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.

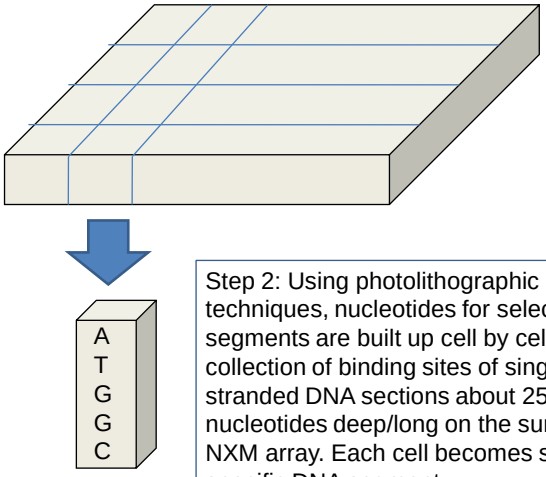
Microarrays is a unique approach which allows for the analysis of millions of samples, it is a marriage of high tech solid state chip technology with DNA bonding. We describe it in the following four steps, each step accompanied by a Figure.

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.

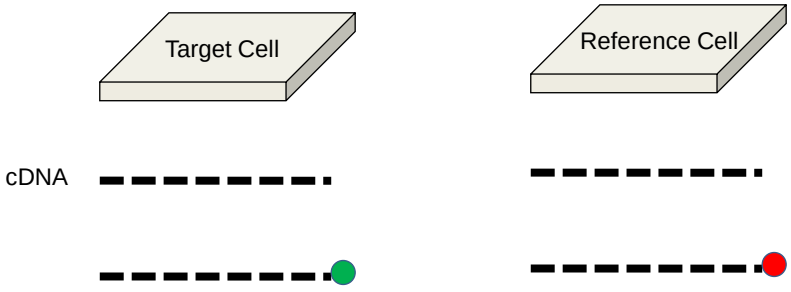
Microarray II



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.

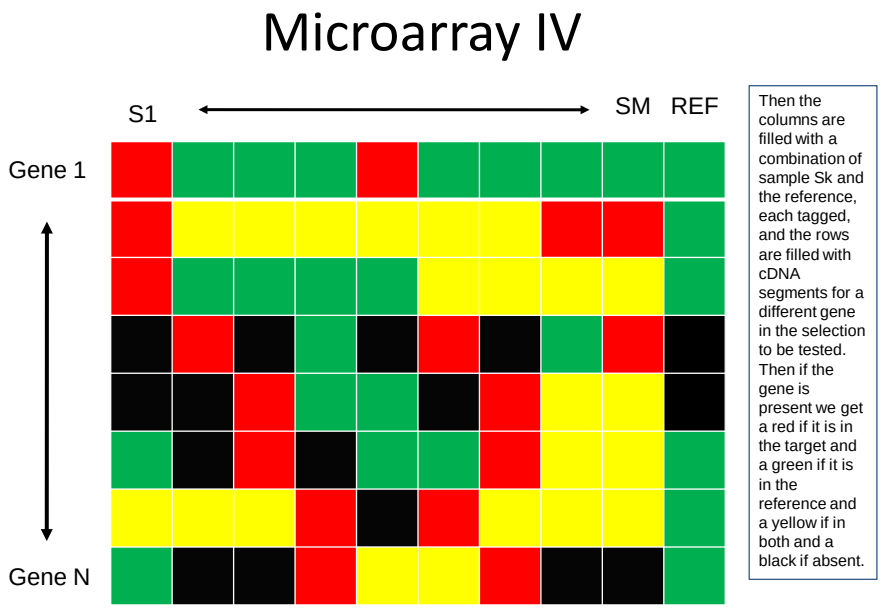
Step 3: Now we take two DNA samples, one from what we call the Target, the plane we wish to categorize. We then take the segments we collected in step one and tag then with green or red tags, green say for the Target and Red for the Reference.

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.

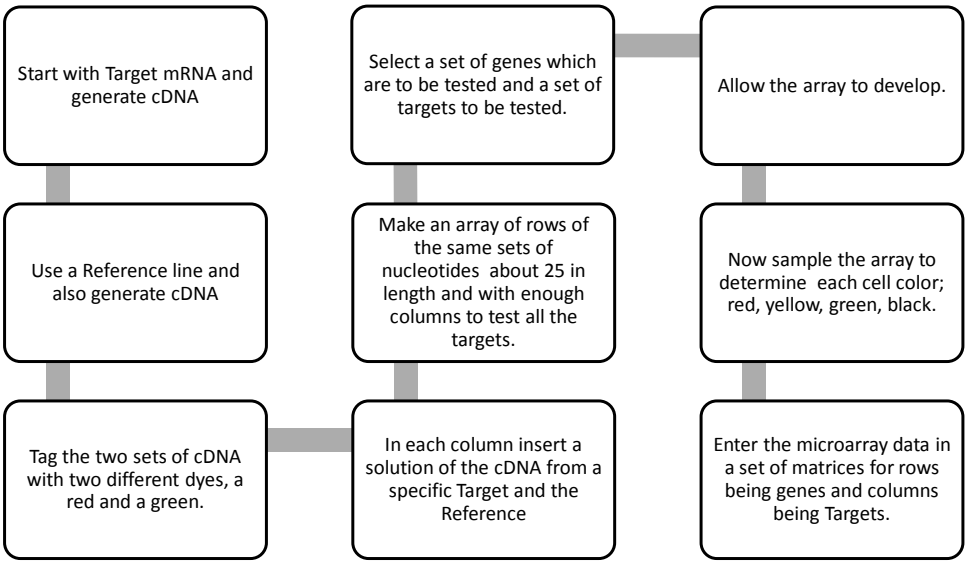
Step 4: We then take the samples from the differing plants, one in each column, and look at the array. If the microarray cell has the gene sequence we are seeking to march, and the Target has that sequence, it will bond and stick. If the Reference has it, it too will bond. If we just get the Target the 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.

5.2.3 Bayesian Network Summary

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.

5.3 STATE DYNAMIC NETWORKS

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.

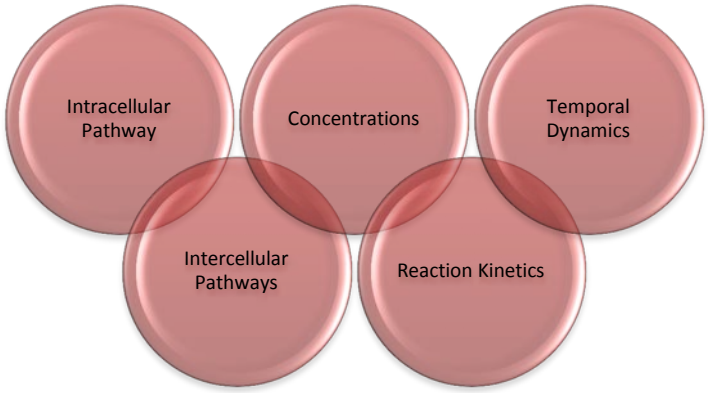
2. Intercellular Pathways: This is the modeling of pathways between and amongst the cells. It is intercellular signalling pathways where receptors and their associated ligands are studied. Again this may generally be focused on concentrations of ligands and similar extracellular and intercellular communicators.

3. Concentrations: The focus is on concentrations of the gene products. 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 are 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.

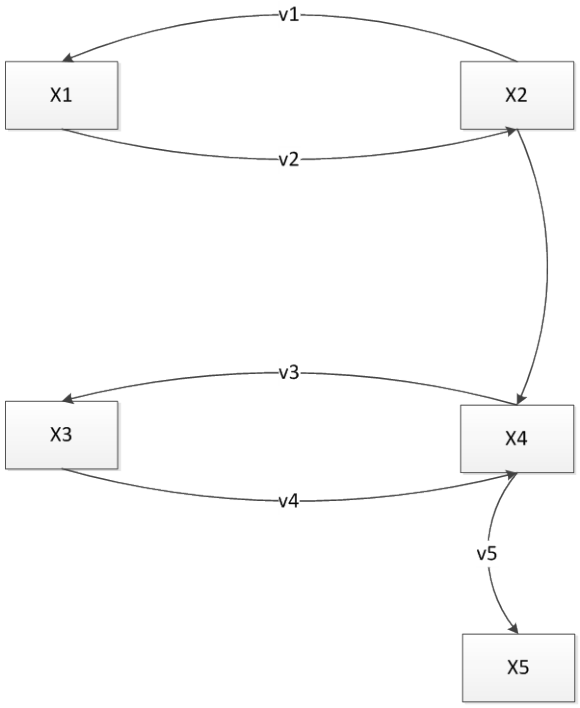
5. Time Variations: The models are all time varying. Thus understanding reaction kinetics from a temporal perspective is essential. This may take our knowledge to the limit and perhaps beyond at the current time.

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.

Let us consider a simple model as below:



Recall that if we have a pathway with say N=5 constituents then we have the following formula 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:

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

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 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.

5.3.1 Kalman Filter

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:

$E[n(t)] = 0$

and

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

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 the concentrations of the respective proteins are given by the 2 x 2 vector as follows:

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

The color model remains the same.

The system model is 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:

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

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

$$g(a,t) = g(a_0,t) + C(a_0,t)\left[a(t) - a_0(t)\right] +$$
$$\frac{1}{2}\sum_{i=1}^{N} \gamma_i \left[a - a_0\right]^T F_i \left[a - a_0\right] + \dots$$

Where we have:

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

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
$$\frac{dP(t)}{dt} = -P(t)C^T(t)K^{-1}C(t)P(t) +$$
$$\sum_{i=1}^{N} PF_iP\gamma_i{}^T K^{-1}(z - g(a_0)$$
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 derived from the Newton method.

5.3.2 The Matched Filter Approach

This is a different approach and it is an application of signal detection taken from classic communication theory. It assumes we have N signals and each signal shape is known but the amplitude of the individual signals is not known. Then we ask how we can estimate the amplitude of each signal if what we have is a received signal which is the sum of the N plus noise. We begin this approach as follows:

Let us assume there are two waveforms bounded on an interval [0, T]

Let

$$s_1(t) = s_1^{orthog}(t) + s_1^{remain}(t)$$

$$s_2(t) = s_2^{orthog}(t) + s_2^{remain}(t)$$

such that

$$\int_0^T s_1^{orthog}(t) s_2^{orthog}(t) dt = 0$$

Now there are three questions which we may pose:

1. Does such a decomposition exist, if so under what terms?

2. What is a constructive way to perform the decomposition?

3. Is there an optimum decomposition such that the "distance between the two orthogonal signals is maximized"?

Namely:

$\exists$ *a set* $\{s_1, s_2\}$

such that

$$\int_0^T s^2{}_1(t)dt = E_1$$

$$\int_0^T s^2{}_2(t)dt = E_2$$

and

$\exists \max$

$$\int_0^T s^{2,orth}{}_1(t)dt = \widetilde{E}_1$$

$$\int_0^T s^{2,orth}{}_2(t)dt = \widetilde{E}_2$$

Let us approach the solution using the theory of orthogonal functions[1]. Now we can specifically use a Fourier series approach. We do the following:

Let

$$s_1(t) = \sum_{n=1}^{\infty} s_1^n \cos(\frac{2\pi}{T} nt) + r_1(t)$$

where

$$r_1(t) = s_1(t) - FS\cos$$

and

$$FS\cos = \sum_{n=1}^{\infty} s_1^n \cos(\frac{2\pi}{T} nt)$$

Likewise

[1] See Sansone, Orthogonal Functions.

Let

$$s_2(t) = \sum_{n=1}^{\infty} s_2^n \sin(\frac{2\pi}{T} nt) + r_2(t)$$

where

$$r_2(t) = s_2(t) - FS\sin$$

and

$$FS\sin = \sum_{n=1}^{\infty} s_2^n \sin(\frac{2\pi}{T} nt)$$

Clearly FS_{cos} and FS_{sin} are orthogonal. The residual functions r are the sin and cos elements respectively of the expansions. We could have just as easily transposed the sin and cos allocations between the two s functions. As to answering the third question we are effectively asking if the r residual functions can be minimized. The answer is not with a Fourier Transform. Then the question would be; is there another set of orthogonal functions which would minimize the residuals, namely:

$$\int_0^T r_1^2(t)dt = R_1$$

and

$$\int_0^T r_2^2(t)dt = R_2$$

are to be minimized. For a Fourier Transform as the orthogonal base we are left with residuals, R, at whatever they may be. However using the Fourier Transform approach we can extract the two signals as follows:

$$P(\lambda) = [C_1]\tilde{s}_1(\lambda) + [C_2]\tilde{s}_2(\lambda) + r_{TOT}(\lambda)$$

Note we can interchange t and λ since they represent the same variable. We now have a "signal" with amplitudes to be determined and a bias which is known. Using standard "signal detection theory" we can readily solve this problem as well. This becomes the "matched filter problem"[2].

What we have sought to accomplish in this paper is to describe color and it generation in plants and to present a set of methods and means to determine the constituents which give rise to those colors. In effect we have created a world view of color, apart from the classic colorimetry approach, and used this and the physical measurements related thereto to affect a method and means to determine concentrations of colorants in flowers.

[2] See VanTrees, Detection, Estimation and Modulation Theory. He presents details on this solution.

The simple application of Beer's law and the use of the known spectra of the proteins allow us to use data from FTS to determine the concentrations of each colorant on literally a cell by cell basis. Beer's law is a simplistic but fairly accurate and consistent method. It would be interesting to explore the details of the transmission of light to a deeper level but the complexity of the cell structure prohibits that at this time.

Having a methodology of the type developed herein we can now more readily examine the genetic pathways and expression. This paper details multiple ways to ascertain concentrations on a cell by cell basis.

6 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. 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.

6.1 TOTAL CELLULAR DYNAMICS MODELS

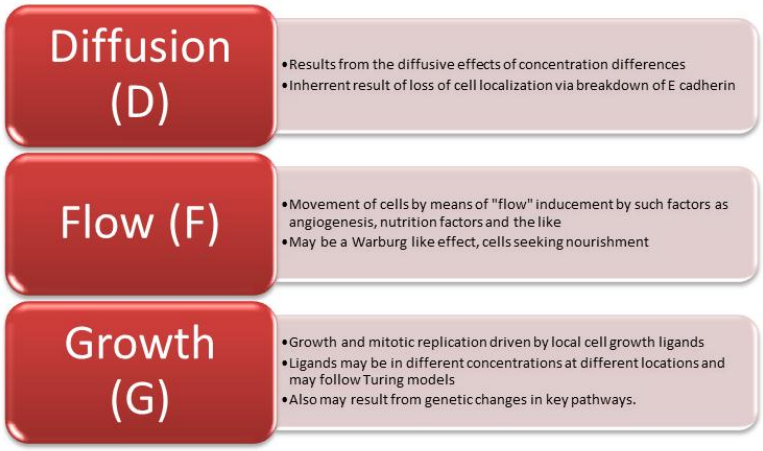
Now depending on the case we assume we can model the flow of cancer cells according to some simple dynamic distributed models[3]. Thus we could have:

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

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.

Now this description has certain physical realities.

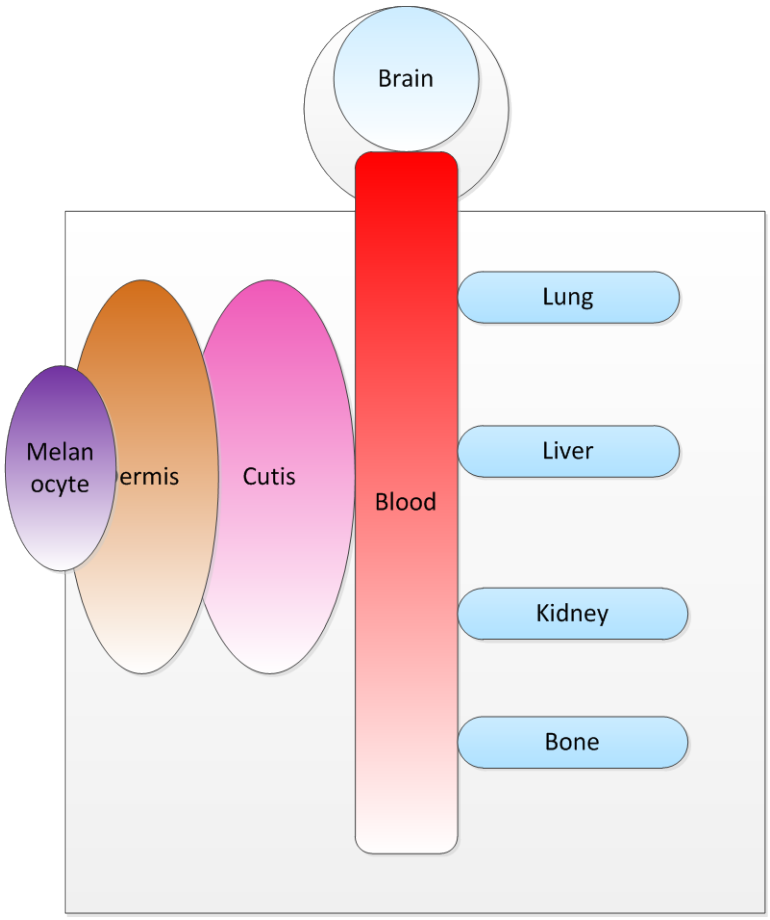
[3] See Andersen p 277 of Bellomo et al for an variant on what we are proposing here. The Andersen model is somewhat similar but lacks the detail we present herein. Also there is in the same volume a paper by Pepper and Lolas focusing on the dynamics of the lymphatic cancer system, p 255.



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,

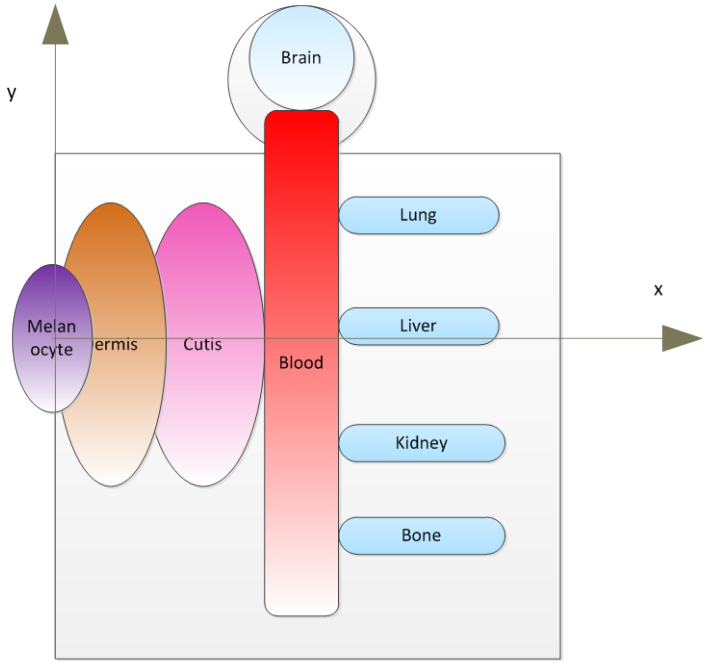


We have a D, E, F, for each gross portion of the body. We also have a model as specifically below in the Table:

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 are 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.

6.2 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.

6.3 DETERMINING THE COEFFICIENTS

The question now is how does one 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:

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

where the constants are to be determined. 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 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 suppositional. But we 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:

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

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

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

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.

7 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

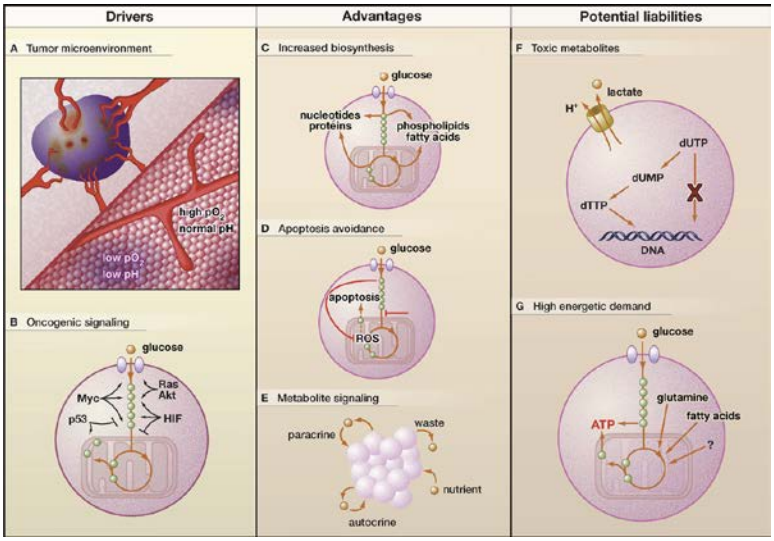
Although the Hypothesis was and is consistent with observations the cause and effect are highly suspect. Yet it does pose the issue of energy infusion in cancer cells.

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.

They demonstrate this in the Figure below showing pathway elements and advantages and disadvantages of such signalling:



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 comments 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 the preference of 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.

8 STOCHASTIC MODELS

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.

9 STABILITY

With highly complex systems, even just linear ones and more so with nonlinear ones we have significant issues regarding 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.

10 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

10.1 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.

10.2 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.

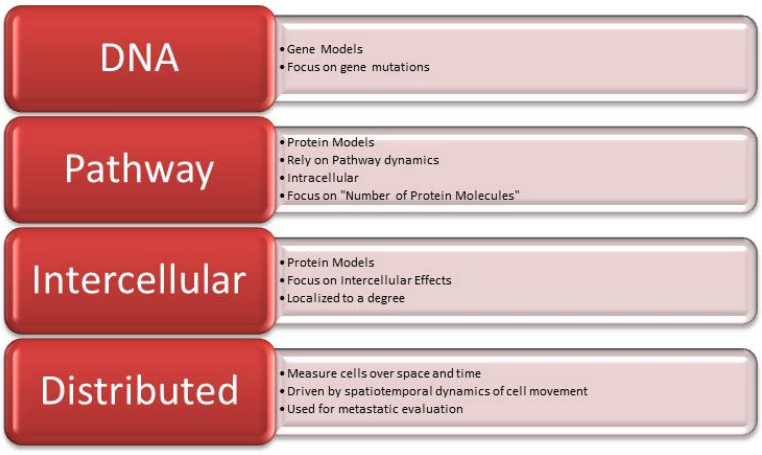
Thus with these two concepts we have a model for a system which can be used to both identify all states and drive it to a desired state.

11 SUMMARY

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 is 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.

12 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. Goentoro, L., et al, The Incoherent Feedforward Loop Can Provide Fold Change Detection in Gene Regulation, Mol Cell 2009 pp 894-899.

49. Goldberg, A., Protein degradation and protection against misfolded or damaged proteins, NATURE, Vol 426, 18/25 December 2003.

50. Gottwein, E., B. Cullen, Viral and Cellular MicroRNAs as Determinants of Viral Pathogenesis and Immunity, Cell Host 2008, pp 375-387.

51. Griffiths, A., et al, **Genetic Analysis 5th Ed**, Freeman (New York) 1993.

52. Grochow, J., M.. Kellis, Network Motif Discovery Using Subgraph Enumeration and Symmetry Breaking, Speed and Huang Ed, RECOMB Springer 2007 pp 92-106.

53. Hanley, J., B., McNeil, The Meaning and Use of the Area under a Receiver Operating Curve, Radiology 1982 V 143, pp 29-36.

54. Hatzimanikatis, V., Dynamical Analysis of Gene Networks Requires Both mRNA and Protein Expression Information, Metabolic Engr, Vol 1, 1999, pp. 275-281.

55. Hay, N., The Akt-mTOR tango and its relevance to Cancer, Cancer Cell, V 8, 2005, pp 179-183.

56. He, L., G. Hannon, MicroRNAs: Small RNAs with a Big Role in Gene Regulation, Nature, 2004 pp 522-532.

57. Herman, J., S. Baylin, Gene Silencing in Cancer in Association with Promoter Hypermethylation, NEJM 2003, pp 2042-2034.

58. Holmes, S., Graphs and Networks in Bioinformatics, Stanford Univ Notes.

59. Hosmer, D., S., Lemeshow, Applied Logistic Regression, Wiley (New York) 2000.

60. Houston, P., Chemical Kinetics and Reaction Dynamics, Dover (Mineola) 2001.

61. Hsu. P, David M. Sabatini1,2,3 Cancer Cell Metabolism: Warburg and Beyond, Cell 134, September 5, 2008.

62. Huang, C., et al, A Study of Cancer Related MicroRNAs through Expression Data and Literature Search, World Acad Sci & Engr, 2009 254-256.

63. 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.

64. 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.

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

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

67. 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

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

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

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

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

72. 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.

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

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

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

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

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

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

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

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

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

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

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

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

85. 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

117. 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.

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

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

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

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

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

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

124. 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.

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

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

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

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

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

130. Perkins, T., et al, Inferring Models of Gene Expression Dynamics, Journal of Theoretical Biology, Vol 230, 2004, pp. 289-299.

131. 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.

132. 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- .

133. Punglia, R., et al, Effect of Verification Bias on Screening for Prostate Cancer, NEJM, 2003 V 349, pp 335-342.

134. Putzi, M., A. De Marzo, Morphologic Transitions Between Proliferative Inflammatory Atrophy and High Grade Prostatic Intraepithelial Neoplasia, Adult Uro 2000 pp 828-832.

135. 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.

136. Reya, T., et al, Stem Cells, Cancer, and Cancer Stem Cells, Nature, 2001 pp 105-111.

137. Robertson, K., A. Wolffe, DNA Methylation in Health and Disease, Nat Rev Gen 2000 pp 11-19.

138. Robertson, K., DNA Methylation, methyltransferase and cancer, Oncogene 2001 pp 3139-3155.

139. Rosen, J., C. Jordan, The Increasing Complexity of the Cancer Stem Cell Paradigm, Science, 2009, pp 1670-1672.

140. Rosenberg, S., Gene Therapy for Cancer, JAMA 1992 pp 2416-2419.

141. Rosenberg, S., The Transformed Cell, Putnam (New York) 1992.

142. 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.

143. 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.

144. Rubinow, S., Mathematical Biology, Dover (New York) 2002.

145. Sachs, K., et al, Causal Protein Signalling Networks Derived from Multiparamemeter Single Cell Data, Science, 2005, pp 523-529.

146. Sage A. P., J. Melsa, System Identification, Academic Press (New York) 1971.

147. Saitou, N., M. Nei, *The Neighbor Joining Method: A New Method for Reconstructing Phylogenetic Trees*, Molecular Biological Evolution, Vol 4 1987 pp 406-425.

148. Schnell. S, T. Turner, Reaction Kinetics in Intracellular Environments with Macromolecular Crowding, Biophys and Molec Bio vol 85 2004 pp. 235-260.

149. Segel, L., M. Slemrod, The Quasi Steady State Assumption: A Case Study in Perturbation, SIAM, 1989, pp 446-477.

150. Seoane, J., et al, Integration of Smad and Forkhead Pathways in the Control of Neuroepithelial and Glioblastoma Cell Proliferation, Cell 2004, pp. 211-223.

151. Setlur, S., et al, Integrative Microarray Analysis of Pathways Dysregulated in Metastatic Prostate Cancer, Cancer Res, V 67, 2007, pp 10296-10303.

152. Sevigani, C., et al, Mammalian microRNAs: A Small World for Fine-tuning Gene Expression, Mam Gene V 17 2006, pp 189-201.

153. 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.

154. Sharifi, N., et al, Androgen Deprivation Therapy for Prostate Cancer, JAMA, V 294, 2005, pp 238-244.

155. Shmulevich, I, et al, Gene Perturbation and Intervention in Probabilistic Boolean Networks, Bioinfo 2002 pp 1319-1331.

156. Shmulevich, I. et al, Control of Stationary Behavior in Probabilistic Boolean Networks, J Bio Sci 2002, pp 431-445.

157. Shmulevich, I., E., Dougherty, Genomic Signal Processing, Princeton (Princeton) 2007.

158. Shmulevich, I., et al, From Boolean to Probabilistic Networks as Models of Genetic Regulatory Networks, Proc IEEE 2002 pp 1778-1792.

159. Shyam, E. et al, The ERG Gene: A Human Gene Related to the ETS Oncogene, PNAS 1987, pp 6131-6135.

160. Sica, G., Bias in Research Studies, Radiology, 2006 V 238 pp 780-789.

161. Simon, W., Mathematical Techniques for Biology and Medicine, Dover (New York) 1986.

162. Slingerland, J., M. Pagano, Regulation of Cdk Inhibitor p27 and Its Deregulation in Cancer, Jrl Cell Phys 2000 pp 10-17.

163. Spirtes, P., et al, Constructing Bayesian Network Models of Gene Expression Networks from Microarray Data, .

164. Stein, G., et al, The Molecular Basis of Cell Cycle and Growth Control, Wiley (New York) 1999.

165. Strathdee, G., R., Brown, Aberrant DNA Methylation in Cancer; Potential Clinical Interventions, Exp Rev Mol Med, 2002, pp 1-17.

166. Studier, J., K. Kappler, *A Note on the Neighbor Joining Algorithm*, Molecular Biological Evolution, Vol 5 1988 pp 729-731.

167. Styerberg, E., et al, Prediction of Indolent Prostate Cancer, Jrl Uro, 2007, V 177, pp 107-112.

168. Su, L., Early Diagnosis and Treatment of Cancer: Prostate Cancer, Saunders(New York) 2010.

169. Swanson, K, et al, Prostate Specific Antigen, Am Jrl Clin Path, 2006, V 125 pp 331-333.

170. Szallasi, Z. System Modeling in Cellular Biology: From Concepts to Nuts and Bolts. MIT Press (Cambridge) 2006.

171. Taby, R., J-P Issa, Cancer Epigenetics, CA 2010 pp376-392.

172. Tang, D., et al, Prostate Cancer Stem/Progenitor Cells, Mol Car 2007 pp. 1-14.

173. Taylor, B., Integrative Genomic Profiling of Human Prostate Cancer, Cancer Cell, V 18, 2010, pp 1-12.

174. Theodoridis, S., K., Koutroumbas, Pattern Recognition, Academic (New York) 2009.

175. Turing, A., *The Chemical Basis of Morphogenesis*, Phil Trans Royal Soc London B337 pp 37-72, 19459.

176. van der Heide, L., et al, The Ins and Outs of FOXO Shuttling, Biochem Jrl 2004, pp 297-309.

177. Vander Heiden, M., et al, Understanding the Warburg Effect, Science 2009 pp 1029-1033.

178. Vander Heiden, M., et al, Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation, SCIENCE VOL 324 22 MAY 2009.

179. Vazquez, F., et al, The PTEN Tumor Suppressor Protein, Biochimica et Biophysica, 2000 pp 21-25.

180. 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.

181. 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.

182. Vliestra, R., et al, Frequent Inactivation of PTEN in Prostate Cancer Cell Lines and Xenografts, Can Res, V 58, 1998, pp 2720-2723.

183. Vogelstein, B., K., Kinzler, Cancer Genes and the Pathways They Control, Nature Med, V 10, 2004, pp 789-799.

184. Vohradsky, J., Neural Network Model of Gene Expression, FASEB Journal, Vol 15, March 2001, pp. 846-854.

185. Wang, E., Cancer Systems Biology, CRC (Boca Raton) 2010.

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.