CYTOGENETIC BIOINFORMATICS OF CHROMOSOMAL ABERRATIONS AND GENETIC DISORDERS: DATA-MINING OF RELEVANT BIOSTATISTICAL FEATURES

by

Jagadeshwari Karri

A Thesis Submitted to the Faculty of

the College of Engineering and Computer Science

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

Florida Atlantic University

Boca Raton, Florida

December 2012

CYTOGENETIC BIOINFORMATICS OF CHROMOSOMAL ABERRATIONS AND GENETIC DISORDERS: DATA-MINING OF RELEVANT BIOSTATISTICAL FEATURES

by

Jagadeshwari Karri

This thesis was prepared under the direction of the candidate's thesis advisor, Dr. Perambur Neelakanta, Department of Computer & Electrical Engineering and Computer Science, and has been approved by the members of her supervisory committee. It was submitted to the faculty of the College of Engineering and Computer Science and was accepted in partial fulfillment of the requirements for the degree of Master of Science.

SUPERVISORY COMMITTEE:

Perambur Neelakanta, Ph.D.
Thesis Advisor

Abhijit Pandya, Ph.D.

Mirjana Pavlovic, Ph.D.

Dolores De Groff, Ph.D.

Borko Furht, Ph.D.
Chair, Department of Computer & Electrical Engineering and Computer Science

Mohammad Ilyas, Ph.D., P.E.
Interim Dean, College of Engineering and Computer Science

Barry T. Rosson, Ph.D.
Dean, Graduate College

November 14, 2012
Date

ACKNOWLEDGMENTS

This study would not have been completed without the direction and guidance from Dr. Perambur S. Neelakanta, the Committee Chairperson and thesis advisor. Dr. Neelakanta taught me how to be patient as well as he was available to provide constructive feedback regarding questions related to research, algorithms and computations. My sincere thanks are extended to him for his time and editing this thesis.

I wish to thank the Committee Members, Drs. Abhijit Pandya, Mirjana Pavlovic and Dolores DeGroff for their time serving as members of the committee and providing their educated assessment to this thesis.

Finally, I would like to greatly thankful to my parents, my in-law and my husband and son for their moral support, constant encouragement and enormous patience while preparing this thesis and especially for all those years of pursuing my education.

ABSTRACT

Author: Jagadeshwari Karri

Title: Cytogenetic Bioinformatics of Chromosomal Aberrations and Genetic Disorders: Data-mining of Relevant Biostatistical Features

Institution: Florida Atlantic University

Thesis Advisor: Dr. Perambur S. Neelakanta

Degree: Master of Science

Year 2012

Cytogenetics is a study on the genetic considerations associated with structural and functional aspects of the cells with reference to chromosomal inclusions. Chromosomes are structures within the cells containing body's information in the form of strings of DNA. When atypical version or structural abnormality in one or more chromosomes prevails, it is defined as *chromosomal aberrations* (CA) depicting certain genetic pathogeny (known as *genetic disorders*). The present study assumes the presence of normal and abnormal chromosomal sets in varying proportions in the cytogenetic complex; and, stochastical mixture theory is invoked to ascertain the information redundancy as a function of fractional abnormal chromosome population. This bioinformatic measure of redundancy is indicated as a track-parameter towards the progression of genetic disorder, for example, the growth of cancer. Lastly, using the

results obtained, conclusions are enumerated, inferences are outlined and directions for future studies are suggested.

CYTOGENETIC BIOINFORMATICS OF CHROMOSOMAL ABERRATIONS AND GENETIC DISORDERS: DATA-MINING OF RELEVANT BIOSTATISTICAL FEATURES

LIST OF FIGURES x
LIST OF TABLES xii
CHAPTER I 1
INTRODUCTION 1
1.1 General 1
1.2 Cytogenetics: A Review 2
1.3 Chromosomes: An Outline 2
1.4 Chromosomal Aberrations 3
1.5 Scope and Objectives of the Present Study 6
1.6 Motivation and Contributions 7
1.6.1 Summary of Contributions 8
1.7 Thesis Layout 9
1.8 Closure 10
CHAPTER II 11
CYTOGENETIC COMPLEX SYSTEM 11
2.1 Introduction 11
2.2 Cellular Constituents: Chromosomal Complexity 11
2.2.1 Centromere 12
2.2.2 Complexity of Eukaryotic Genomes 19
2.2.3 Types of Centromere Sequences 19
2.2.4 Wet-lab characterization of the Centromere 20
2.3 Centromere Aberrations: Disorganization of Cytogenetic Framework 21
2.4 Chromatin - Structural and Functional Modifications: A Review 22

2.4.1 Structure of DNA and Chromatin 23
2.4.2 Chromatin Organization.....27
2.4.3 Abnormal Features of Chromatin 29
2.5 Closure 30
CHAPTER III 31
CHROMOSOMAL ABERRATIONS 31
3.1 Introduction.....31
3.2 Chromosomal Aberrations: An overview 31
3.3 Features of Chromosomal Aberrations 34
3.4 Chromosomal Aberrations and Genetic Disorders 37
3.5 Chromosomal Aberrations: Oncological Considerations 37
3.6 Closure 41
CHAPTER IV 42
MIXTURE OF NORMAL AND ABNORMAL CHROMOSOMES: EVALUATION OF CHROMOSOMAL ABNORMALITY VIA SIMPLE PROPORTION MIXTURE MODEL ...……...42
4.1 Introduction.....42
4.2 Cytogenetic Mixture Contents: Statistical Implications 45
4.3 A Case Study.....45
4.3.1 Case Study Example: Maternal Cell Contamination 46
4.3.2 Computational Details and Discussion on Case Study Example.....47
4.4 Results.....49
4.4.1 Case-1a: "Y+" Spontaneous Abortions.....49
4.4.2 Case-1b: "Y−" Spontaneous Abortions 51
4.5 Discussion and Closure.....51
CHAPTER V .....…..53
STATISTICAL MIXTURE MODEL OF NORMAL AND ABNORMAL CHROMOSOMAL ADMIXTURE: APPLICATION TO CANCER GROWTH MODELS .....……53
5.1 Introduction.....……53

5.1.1 Proportional-content Theoretics Applied to Cytogenetic Constituents: A Revisit ……..……..……..54
5.2 A Review on Statistical Mixture Theory: Applications to Cytogenetic Context……...…..55
5.2.1 Informational efficiency (η) pertinent to Cytogenetic Informatics...…….55
5.2.2 Quantifying the Mixture Attributes of Cytogenetic Complexity via Redundancy factor……...……..…57
5.3 Statisitcal Mixture of Cytogenetic Contents: Implications on Observed Pathogenic States ……...……60
5.4 Genetic Disorder and Cancer-Growth Considerations: Modeling *via* Copy number Altrations…….……...……...….60
5.5 Ploidy and Aneuploidy Involvement in Tumor Growth……..……...62
5.6 A Mixture-state Model of Cytogenetic Complex in Terms of Clonal and Non-clonal Alterations……62
5.7 Dynamics of CCA and NCCA Profiles……..…….....64
5.7.1 Quantifying Stochastical Mixture Attributes of the Contents in the Cytogenetic Complexity ……...……64
5.8 Temporal Changes in the Cytogenetic Constituents : Neoplasmic evolution……..……..……..…...69
5.8.1 Chromosomal Aberrations: A Stochastical Profile…….……...71
5.8.2 Stochastical Dynamics and Bernoulli-Riccati Equation on Tumor Growth Model ……..…….....71
5.9 Cancer Growth Models...……..…….....74
5.10 Discussions and Concluding Remarks……...……..…85
CHAPTER VI 87
RESULTS, DISCUSSIONS AND CONCLUSIONS…….....87
6.1 Introduction.....87
6.2 Discussion and Inferences……..……..89
6.3 Open-questions for Future Study……90
6.4 Closure 90

REFERENCES……103

LIST OF FIGURES

Figure 1.1: Chromosome structure with its components, DNA and histones…… 4

Figure 1.2: Single chromosome mutation: Deletion, Duplication and Inversion….……...5

Figure 1.3: Two chromosome mutations: Insertion and Translocation……...5

Figure 2.1: Diagrammatic representation of a centromere…….....12

Figure 2.2: Diagrammatic represtation of cell division namely mitosis and meosis ……14

Figure 2.3: Structure of chromatin…….25

Figure 3.1: Oncogenic tree model……..41

Figure 4.1(a): A risk-factor versus an ensemble of trails (T_1, T_2...T_{14}) of data presented for Table 4.1 for (Y+) state…….….49

Figure 4.1(b): A risk-factor versus an ensemble of trails (T_1, T_2...T_{14}) of data presented for Table 4.1 for (Y−) state…….….50

Figure 5.1: Proposed pathway for cancer evolution and progression when kn is equal to the pseudo reaction rate constant …..…….….64

Figure 5.2: A complex system **Ω: X** depicting a mosaic of statistical mixture constituted by a pair of binary subsystems (compositional domains)…….65

Figure 5.3: Plot of $L_Q(\tau)$ and $1 - L_Q(\tau)$ of Table 5.3……..……...77

Figure 5.4: Experimental data in normalized form of C(τ), N(τ) and CH(τ) and τ denotes the stage of cancer growth as available in [5.9]…….….79

Figure 5.5: Plots of C(τ), N(τ) and CH(τ)…….…….....80

Figure 5.6(a): Growth function G(t) with respect to normalized time t_N …….……...82

Figure 5.6(b): Decay function D(t) with respect to normalized time t_N …….……83

Figure 5.7(a): Growth function G(t) with respect to normalized time t_N …….……...84

Figure 5.7(b): Decay function D(t) with respect to normalized time t_N …….……85

LIST OF TABLES

Figure 4.1: Possible maternal cell contamination details in two groups of “46,XX” spontaneous abortions with or without Y chromosome[4.1]…….43

Figure 4.2: Proportional Modeling the structure of cytogenetic factors of prenatal selection corrected for cell contamination [4.1]……..…….48

Table 5.1: Stage specific mutation pseudo reaction rate constants [5.9] …….……74

Table 5.2: Calculation of C*, N* and CH*…….…….……75

Figure 5.3: Experimental and empirical data on N(τ) and C(τ) of [5.9] and computed data of the present model …….……76

CHAPTER I

INTRODUCTION

1.1 General

Seeking a statistical methodology in bioinformatics to analyze chromosomal aberrations at cellular level forms the primary goal of the present study. As well known, chromosomal aberration could be a major cause of human diseases implicated by the so-called *genetic disorders*. That is, while the collection of normal chromosomes in the cells depict a healthy cytogenetic status, the presence of any abnormal chromosomal details would imply atypical conditions within the cell reflected as possible genetic disorders.

If the cytogenetic complex is devoid of any aberrations, its associated features that decide the normal cellular functions are largely non-random and mostly deterministic; as such, the cell at cytogenetic level can be considered as ‘organized’ and entropy-free. On the other hand, should this complex contain aberrated versions of chromosomes, it can be regarded as being ‘disorganized’; and, the associated extent of disorganization can be modeled in terms of the accompanying entropy details. Such entropy aspects of abnormal features cells would determine the extent of the progression of the associated disease (genetic disorder). Typically, cancer for example, is viewed in terms of unregulated cell growth promoted by chromosomal aberrations.

In short, the present study is objectively indicated to pursue relevant studies in analyzing the extent of abnormal features at cytogenetic complex in terms of the associated

dis-organizational statistics and entropy characteristics. Relevant informatics of the cytogentic framework forms a new and novel branch of bioinformatic tasks useful in genetic disorder studies.

1.2 Cytogenetics: A Review

Cytogenetics is a branch of genetics that correlates the structure, number and behavior of chromosomes with hereditary and certain diseases. It is a study concerned with cells and its constituent *chromosomes*. Chromosomes basically depict a collection DNA strings seen within the cells. They can be microscopically examined and characterized by a process known as *karyotyping*, which is an effort to elucidate the number and structure of chromosomes. Chromosomal features, when seen distorted, denote abnormalities that can be identified in a diverse spectrum of disease states, particularly in humans. For example, early embryonic death, minor-to-major congenital defects, development of cancer and infertility (or sterility) can be mentioned as those due to genetic disorders caused by chromosomal aberrations. Relevantly, informational entropy can be adopted to quantify the extent of such chromosomal aberrations at the cellular level. The underlying efforts form the theme of the present research.

1.3 Chromosomes: An Outline

Normal human-body consists of about 50 trillion cells. In the nucleus of each cell, genetic material is well-organized with compactly packaged DNA and histone proteins. The plethora of such structures is called *chromosomes*. DNA and histones bind together due to the existence of electrostatic forces between negatively-charged phosphate group in the DNA and positively-charged amino acids (AAs) in the histone proteins. Structurally, chromosomes may vary widely across different organisms; in general, DNA

molecule could be circular or linear. Eukaryotic cells have large linear chromosomes while prokaryotic cells have small circular chromosomes. Further, cells might contain more than one type of chromosome in an organism, for example, mitochondria in eukaryotes and chloroplast in plants.

Somatic cell of humans consists of 46 chromosomes organized into 23 pairs. In each pair, one chromosome comes from maternal and the other from paternal source. There are 22 pairs of *autosomes* that determine the genetic traits and one pair of *allosomes* that specifies sex, typically indicated as XX for female, and XY for male [1.1].

Each chromosome has a constriction point known as *centromere*, which divides the chromosome into two sections or arms. The short-arm of the chromosome is labeled as *p-arm* and the long-arm of the chromosome is labeled as *q-arm*. The location of the centromere gives the chromosome its characteristic shape.

Chromosomes are essential units responsible for cellular division and they must be replicated, divided, and passed successfully to their daughter cells so as to ensure genetic diversity and survival of the progeny. Moreover, an ordered organization of the genetic material at molecular level is required by the cell for its normal functionality across all living systems. A typical chromosomal structure is illustrated in Figure 1.1.

1.4 Chromosomal Aberrations

Chromosomal aberrations reflect the abnormality aspects of the chromosome *vis-à-vis* the number or structure. They usually occur when there is an error in cell-division following *meiosis* or *mitosis.* Some factors that influence chromosomal aberrations are as follows: Increased maternal age, abiotic environment, ionizing radiations, autoimmunity, viral infections and chemical toxins etc.

Numerical aberrations in chromosomes occur when a default is encountered in the usual number of chromosomal pairs. When an individual is missing a pair of chromosomes, a condition known as *monosomy*, as in Turners syndrome 45X occurs; or, when an individual has more than two chromosomes in a pair, a genetic disorder occurs as observed in *Downs syndrome 21 Trisomy*.

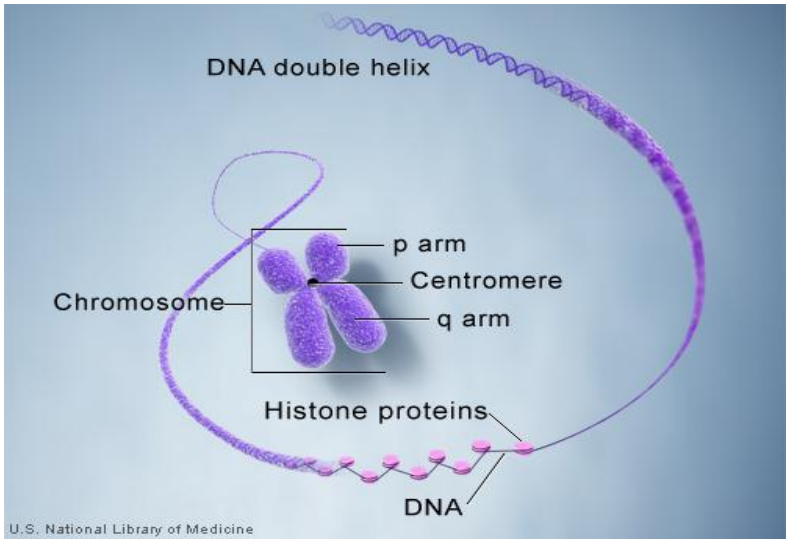

Figure 1.1: Chromosome structure with its components, DNA and histones.

(Adapted from US National Library of Medicine, Available at: http://ghr.nlm.nih.gov/handbook/illustrations/chromosomestructure)

Structural aberrations in chromosomes imply whenever an individual chromosomal structure is altered. Such structural variations can be classified into five categories: *Deletions* depicting a portion of chromosome missing or deleted; *duplications* where a part of chromosome is duplicated resulting in extra genetic material; *translocations* implying a part of chromosome transferred to another chromosome; *inversions* denoting a part of chromosome broken and again getting attached upside down rendering, the genetic material inverted and formation of *rings*, with a portion of chromosome broken off and assuming a

shape of ring, (with or without the loss of genetic material) [1.2]. Typical chromosomal aberrations are illustrated in Figure 1.2 and 1.3.

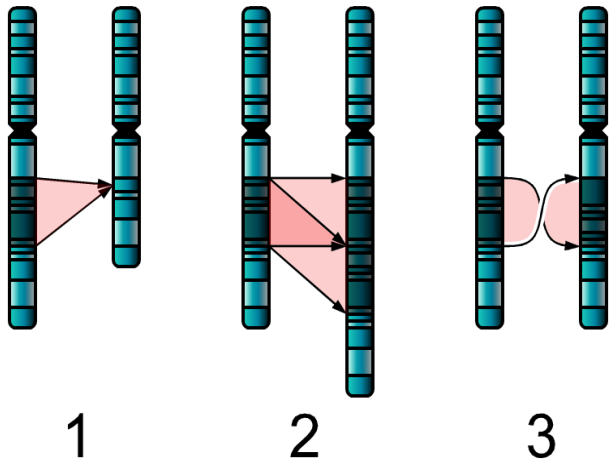

Figure 1.2: Single chromosome mutation (1) Deletion, (2) Duplication and (3) Inversion.

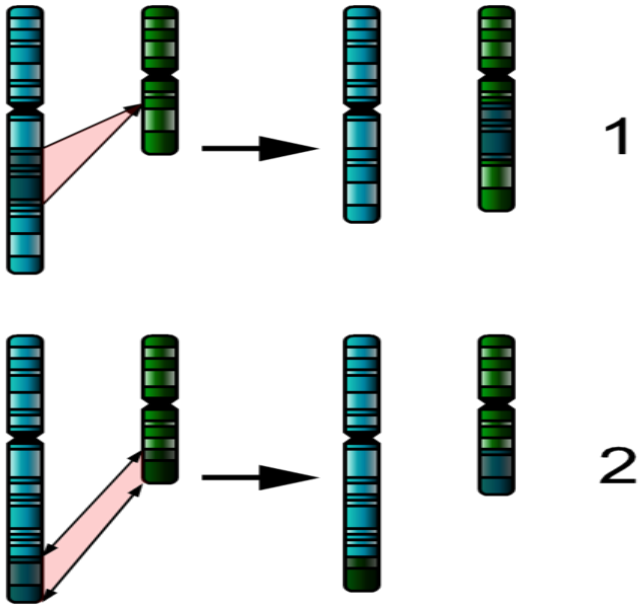

Figure 1.3: Two chromosome mutations: (1) Insertion and (2) Translocation.

(Adapted from Wikipedia, Available at: http://commons.wikimedia.org/wiki/File:Two_Chromosome_Mutations.png and http://commons.wikimedia.org/wiki/File:Single_Chromosome_Mutations.png)

In general, all species can be affected by chromosomal diseases. Prevalent in humans are genetically- defined diseases due to errors made in chromosomal segregation during meiosis or as a result of exogenous influences. As indicated earlier, complications such as minor-to-minor congenital defects, development of cancer, infertility/sterility and early embryonic death etc. are caused by chromosomal abnormality. In view of the wide prevalence of chromosomal disorders and related pathology, it is imperative to acquire a broad-base of relevant knowledge towards proper medical diagnosis *vis-à-vis* genetic disorders. As such, the present study is motivated with a scope and objectives to evolve some research considerations as outlined in the following sections:

1.5 Scope and Objectives of the Present Study

The gross efforts exercised in this research can be summarized as enumerated below to indicate the tasks performed and essential outcomes/contributions derived thereof:

- Understanding the stochastical disorganization at cytogenetic level
- Developing cytogenetics-inspired approach to assess the informational redundancy that results from the admixture of abnormal chromosomes and normal chromosomes
- Deducing relevant algorithms in terms of entropy features of the chromosomal aberrations *via* statistical mixture theory
- Applying entropy-based measures to determine the severity of genetic disorders such as growth of cancer.

The efforts as above are novel in the sense that the information-theoretic approach pursued is unique and new in the context of cellular cytogeny. To the best of author's

knowledge no such cohesive entropy-specific techniques have been exercised in the context of chromosomal aberrations evolving in a cytogenetic system of cellular complexity. Matching the scope of the research conceived, the objectives of the research pursued can be enumerated as follows:

- To survey and compile literature on chromosomal aberrations and related genetic disorders
- To model the cytogenetic complexity in terms of the associated 'disorganization' caused by chromosomal abnormality coexisting with normal chromosomal contents
- To develop relevant models to portray the abnormal features of chromosomes *via* statistical mixture theory
- To formulate an entropy metric (or information-theoretic measure, in Shannon's sense) to compare the contents of an admixture normal and abnormal chromosomes using stochastical mixture theory
- To elucidate the informational redundancy in terms of entropy features seen in the cytogenetic complex system, so as to differentiate normal and abnormal chromosomal implications.

1.6 Motivation and Contributions

The present study, in short is conceived to find compatible quantitative measures on the severity of chromosomal aberrations on the basis of disorganizational features of cytogenetic complexity. Consistent with this objective, the motivated effort addressed in this research is inspired by the quest to seek certain avenues of unexplored strategies in modeling chromosomal abnormalities observed in a cytogenetic complex system *via* entropy considerations. Hence, the informational entropy (in Shannon's sense) arising from

varying proportions of normal and abnormal chromosomes as a stochastical mixture in a cytogenetic system is examined. Relevant quantitative assessment of genetic disorders *versus* chromosomal aberrations can help deciding the severity of related diseases seen in patients.

Inspired by the need, the motivated research tasks performed have resulted in outcomes, which are summarized below:

1.6.1 Summary of Contributions

- A comprehensive outline on cytogenetics for bioinformatic analysis and data-mining: Bioinformatics at cytogenetic level is sparse
- A complex system depiction of the cytogenetic framework: Again, the complex system description of cytogenetics is rarely done
- Constructing the stochastical structure of the cellular system in terms of normal and abnormal chromosomes *via* the complex system considerations
- Hence, proportional and stochastical mixture models on the population of cytogenetic contents is developed
- Introducing the concepts of entropy and information to determine the stochastical features of the cellular complex with normal and abnormal contents
- Applying stochastically-justifiable logistic function description to cancer growth resulting from cytogenetic abnormality. Model results are compared against available data.

1.7 Thesis Layout

In order to cohesively address the research efforts commensurate with the objectives indicated and present the outcomes thereof, this thesis is written with an organized set of chapters as follows:

- Chapter I: Introduction - This (present) chapter provides an introduction to the topic of research pursued with an indication of relevant scope and objectives; and, the general format of thesis organization is outlined
- Chapter II: Cytogenetic Complex System - Elaborated in this chapter are details pertinent to cytogenetic complex system. An archive of literature is reviewed and salient details on the organized as well as disorganized features of the cellular interior are presented consistent with the topic of interest
- Chapter III: Chromosomal Aberrations - This chapter outlines the general aspects of the algorithms prescribed in knowing the extent of normal and abnormal chromosomal entities present in a cytogenetic complex. Available methods are reviewed and discussed
- Chapter IV: Mixture of Normal and Abnormal Chromosomes: Evaluation of Chromosomal abnormality *via* Simple Proportion Mixture Model- With the conceived objective of developing algorithms to estimate the quantitative profiles of normal and abnormal chromosomes in the cells
- Chapter V: Statistical Mixture Model of Normal and Abnormal Chromosomal Admixture: Application to Cancer Growth Models- This is written to indicate an entropy-based approach that determines the relative proportion of chromosomal

aberrations being present using stochastical mixture theory. Hence, the concept of informational redundancy (in Shannon's sense) is invoked to frame a compatible algorithm. The application of the algorithms developed is demonstrated using relevant examples of real-world data sets. The associated analytical pursuits are described and presented along with algorithmic representations. Computational procedures are outlined in evaluating the informational redundancy of the mixture containing normal and abnormal chromosomes

- Chapter VI: Results, Discussions and Summary - This chapter is written to present the briefing of results due to number of experimental simulations using the test algorithms. Relevance of results obtained are discussed with necessary conclusions.

1.8 Closure

This introduction chapter is written to outline the overall content of the thesis and provides details on the scope of the research, underlying objectives and driving motivation. Further, the thesis organization is specified with a format outline on ensuing chapters.

CHAPTER II

CYTOGENETIC COMPLEX SYSTEM

2.1 Introduction

Cytogenetics is concerned with the study of structure and function associated with the cell at chromosomal level. Implicated by genetic abnormalities and medical considerations the art of cytogenetics become widen subject in modern medical science. With the advent of development in genetics in the early part of the last century, a number of studies were undertaken to identify molecular cytogenetics and the relevant considerations in the related pathogenic condition, both wet lab experiments and theoretical considerations have been advocated to examine chromosome structure and learn about the relationship between chromosome phenotype so as to determine the causes of chromosomal aberrations.

Commensurate with the objective of this thesis in ascertaining the disorganized aspects of cellular/chromosomal frame work, this chapter is written to provide a descriptive and illustrative note on the cytogenetic complex system period. Hence presented are relevant definitions as well as anatomical features of cellular constituents along with the associated functional attributes.

2.2 Cellular Constituents: Chromosomal Complexity

As described in chapter 1, a cellular constituent refers to a collection of the original DNA in transcriptomic form known as the chromosomes. As illustrated in Figure 1.1, the

chromosomal structure contains distinct parts. Described in this section is an outline on each part.

2.2.1 Centromere

Centromere refers to a DNA region found in the vicinity of the middle section of chromosomes, where two sister chromatids are in closest contact. Centromere was described by the German biologist, Walter Flemming in 1880s [2.1] as the "primary constriction" of the chromosome facilitating chromosomal inheritance process. Centromere region stains in the fluorescent labeling less strongly than rest of the chromosome. Centromere also plays key role in the so called kinetochore formation and spindle attachment. Shown in Figure 2.1 is a simple diagrammatic depiction of a centromere.

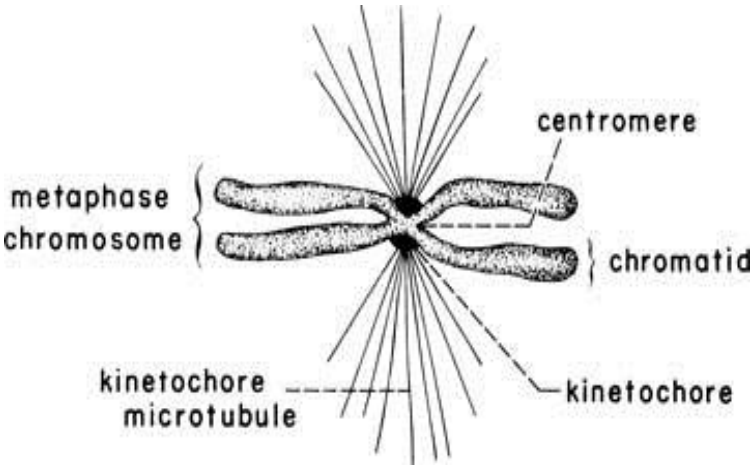

Figure 2.1 Diagrammatic representation of a centromere

(Adapted from Wikipedia, Available at:
http://commons.wikimedia.org/wiki/File:Two_Chromosome_Mutations.png)

With reference to Figure 1.1, it can be seen that a chromosome has two arms p (shorter arm) and q (longer arm). Further, considering the location of the centromere and connection of the arms, the chromosomes can be divided into six types namely, *metacentric, submetacentric, acrocentric, telocentric, subtelocentric and holocentric* (Human's chromosomes are however classified into only four types: Metacentric, submetacentric, acrocentric, and telocentric).

*Metacentric***:** A chromosome is metacentric, if its two arms are approximately equal in length. The formation of metacentric chromosome is due to a balanced translocation or due to the fusion of two acrocentric chromosomes.

Submetacentric: A chromosome is submetacentric if the lengths of the arms are unequal.

*Acrocentric***:** A chromosome is acrocentric, if the p arm is hard to observe in spite of its presence. In humans, chromosome numbers 13,14,15,21 and 22 are acrocentric.

*Telocentric***:** A chromosome is telocentric, if the centromere is located at the terminal end and humans do not have such telocentric chromosomes.

*Subtelocentric***:** A chromosome is subtelocentric, if its centromere is located closer to its end than to its center.

Holocentric: A chromosome is holocentric, if the entire length of the chromosome acts as a centromere [2.1].

The process of cell division can be described in two contexts, namely mitosis and meiosis. The underlying aspects can be understood from Figure 2.2.

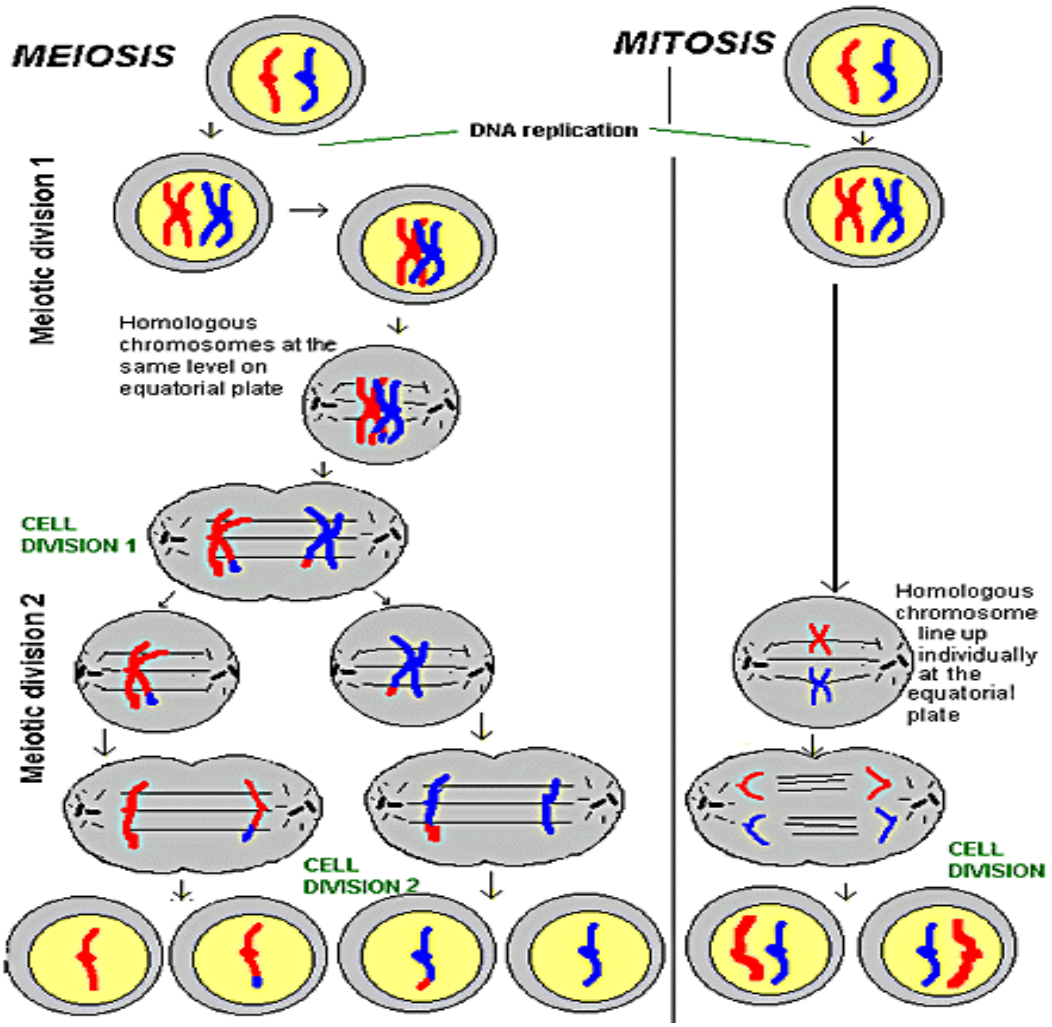

Figure 2.2: Diagrammatic representation of cell division namely, mitosis and meiosis. (Adapted from accessexcellence.org, Available at: http://www.accessexcellence.org/RC/VL/GG/comparison.php)

Centromeres play major role in cell division in mitosis (which divides chromosomes in a cell nucleus) as well as in meiosis (which denotes specific cell division in reproductive phases) by directing accurately relevant segregation of chromosomes. The number of chromosomes is halved in meiosis, with the mitotic division differing from the typical mitotic division in two aspects as follows:

- In first mitotic division, chromosomes from maternal and paternal pair at the beginning of the meiosis, divides into two sister chromatids resulting in four chromosomes
- In mitosis, the sister chromosomes are pulled towards opposite poles, whereas in meiosis four chromosomes attach to a spindle and the sister chromosomes attach to the same pole.

Maternal chromosomes move to one pole and paternal chromosomes move to opposite pole. The separation of maternal and paternal genes during the formation of sperms and eggs is mainly due to centromere's unique role.

In mitosis, sister chromatids are joined together at centromeres until the spindle check point has been passed that happens during metaphase hence resulting in accurate cell division. Centromere dictates the assembly of kinetochore which is defined as a set of proteins that assemble on the centromere providing the point of attachment for the spindle microtubules denoting the part of the cytoskeleton provides structure and shape to a cell, during cell division they participate in the formation of spindle fibers.

Centromeres are regions of specialized chromatin because such regions are conserved during the course of evolution the molecular composition is however can be different in different species, and, centromere identity is preserved during cell division.

Centromere protein- A (also known as CENP-A) remains associated with centromere throughout the cell division as it is the integral component of the nucleosome. In the synthetic phase (or S phase), this CENP-A marks the position of centromere on two DNA strands. CENP-A containing nucleosomes are divided more or less equally between two DNA strands. CENP-A loading factor recruits additional CENP-A to the

centromeres. The activity of CENP-A appears to be fundamental to centromeres, in as much as cells that lack CENP-A fail to divide properly [2.1].

Centromeres consist of highly repetitive DNA regions and bound kinetochore proteins that are proteins are required for the attachment of microtubules to chromosomes during cell division. In eukaryotes, centromere basic function is highly conserved but divergent sequences are observed within closely related species.

Prokaryotes have low copy number of DNA, so that a minimal system would prevail for DNA segregation. Relevant bacteria has plasmids where chromosomes are present and centromere as well as kinetochore-like structures play key role in segregating the chromosomes, in such species *via* partitioning (Par) apparatus that resembles an eukaryotic spindle. Such bacterial par systems encode three elements: (i) Centromere like site on plasmid DNA; (ii) Protein binding to this site and (iii) Actin-like ATPase.

As an example, the par systems of the *Escherichia coli* R1 plasmid contain parC, ParR and ParM and of the *Enterococcus faecium* plasmid pGENT contains cenE, PrgO. Both have similarities such as centromeric regions are curved, these regions consist of two sets of repeats: DNA binding proteins ParR in E.coli and PrgO in *E.faecium* that binds to the centromeric regions of parC in *E.coli* and cenE in *E. faecium* forms nucleoprotein complexes. The segregation proteins ParM and PrgO interact with their corresponding nucleoprotein complexes and form dynamic actin-like filaments for active and directed plasmid partitioning. Centromere's DNA is bendable or curved in order to support tight winding of DNA around protein binding sites from prokaryotes to eukaryotes [2.2].

Budding yeast, *Saccharomyces cerevisiae* the centromeric DNA is very simple in organization, being only of 125 to 200 bp long and this region 125 bp locus known as CEN plays a significant role in mitotic and meiotic chromosome segregation. The CEN sequence is organized into three domains, consisting of two highly conserved protein-binding sites (termed CDE I and CDE III) flanking a 78-86 bp high (A + T) central sequence (CDE II). Mutational analyses have shown that the 25 bp CDE III binding site is absolutely essential for centromere function [2.3].

The centromere functionality is conserved in all eukaryotes from yeast to humans, relevant details are as follows:

- The centromere is the site of kinetochore assembly (the protein complex that drives chromosome segregation); they are formed at one and only one site on each chromosome. Further, centromere depicts the last region where sister chromatids remain tethered by cohesion until *anaphase* (*Anaphase* means third stage of mitosis, where the daughter chromosome move towards opposite poles)
- Centromere incorporates a sensor, known as the spindle checkpoint that monitors attachment of sister kinetochores to microtubules from both poles and hence tension across sister centromeres
- Kinetochore-associated motor proteins are responsible for the movement of chromosomes along microtubules toward the spindle poles.

The centromere of yeast can be studied under two categories namely, (i) budding yeast and (ii) fission yeast. In budding yeast, for example in *Saccharomyces cerevisiae*, it differs from fission yeast or Drosophila melanogaster and humans with the absence of the silent chromatin.

In fission yeast, such as *Schizosaccharomyces pombe*, *Drosophila melanogaster* (also known as fruit fly which belongs to *Drosophilae* family) has basic similarities with humans, such as kinetochores binding to microtubules. The kinetochore is organized in such a way that it coordinates with microtubule binding sites so that movement of the microtubules is well controlled. In yeast, centromeric DNA is closely packed in such a way that the genes placed within those regions are transcriptionally silent. Heterochromatin is typically present in *Drosophila* and humans. Studies on the fission yeast reveal the role of repressive heterochromatin in centromere function.

Human centromeres are relatively large, several million base pairs (bps) and consist predominantly of the same 171-bp sequence, known α-satellite DNA, repeated hundreds of thousands of times and by a number of centromeric proteins (CENPs) . Centromere is recognized by the components of the kinetochore in cell division phases. The centromeric DNA is normally in a heterochromatin state that is condensed state; this is required for recruiting cohesion complex mediated after DNA replication and during anaphase. In this chromatin histone3 is replaced with a centromeric specific variant CENP-A, The presence of CENP-A is believed to be important for the assembly of the kinetochore on the centromere. CENP-C has been shown to localize almost exclusively to these regions of CENP-A associated chromatin. CENP-B, binds specifically to alpha-satellite DNA sequences, CENPs as a result help in regulation of heterochromatic modification, later followed by cell division.

Constitutive proteins are permanently associated with centromere, the key proteins are CENP-A (centromere protein A), CENP- B, CENP-C and CENP-G and the

facultative proteins are associated with centromere temporarily. These are CENP-E, CENP-F and turblin [2.4].

2.2.2 Complexity of Eukaryotic Genomes

Eukaryotic genomes can be regarded as complex systems due to following reasons;

- The introns and exons are such that, the introns are predominant accounting about ten times more than exons
- Eukaryotic genes are normally present in multiple copies known as *gene families* and also they are present as *pseudogenes* (denoting inactivated genes due to mutations) constituting nonfunctional genes
- Repetitive DNA sequences are significantly present in eukaryotic genome. Nearly 40 % is usually repetitive DNA sequences and some of them even prevail as 10^5 to 10^6 copies per genome.

The number of genes in eukaryotes is very high. For example, human genome is estimated to contain about 100,000 genes. Only a small fraction of the genome in the complex eukaryotes corresponds to protein-coding sequences [2.5].

2.2.3 Types of Centromere Sequences

There are two major classes of centromeres, namely, (i) *regional centromeres* wherein the DNA sequences are not defined and the functions of such centromeres are not known. These centromeres contain large amounts of DNA and are often packaged into heterochromatin. They typically consist of large arrays of repetitive DNA (for example, satellite DNA) where the sequence within individual repeat elements is similar, but not identical. As said earlier, the primary centromeres in humans are repeat units

denoted as α-satellite. A number of other sequence types are also found in the regional centromeres for example, as in the fission yeast. (ii) *Point centromeres* denote entities wherein the DNA sequences are both necessary and sufficient to specify the centromere identity as well as the functions in organisms. Point centromeres are smaller and more compact. In budding yeasts, the point centromere region is relatively small, (about 125 bps DNA) and contains two highly conserved DNA sequences that serve as binding sites for essential kinetochore proteins [2.6].

2.2.4 Wet-lab characterization of the Centromere

In fluorescent *in situ* hybridization (FISH), fluorescent probes are used to bind only those parts of the chromosome so that a high degree of sequence similarity can be observed subsequently, *via* fluorescence microscopy to find out where the fluorescent probe is bound to the chromosomes. FISH is used to detect and localize the presence or absence of specific DNA sequences on chromosomes in determining centromere specific satellite sequences that are expected to have the copy number being low.

Immunoprecipitation is another technique where precipitating a protein antigen out of a solution using an antibody (that specifically binds to particular protein) is used to determine human neocentromeres. Such neocentromeres of humans are not composed of repetitive DNA. Hence immunoprecipitation is adopted. A related version namely *chromatin immunoprecipitation* (ChIP) is used with specific antibody followed by pyro-sequencing to determine the order of nucleotides in DNA.

Bacterial artificial chromosome (BACs) is an approach used to sequence the genome of organisms, where a short piece of DNA of the organisms is amplified as an insert in BACs and then sequenced [2.7].

In the wet-lab based centromere DNA sequencing, the associated limitations are as follows: (a) Since centromere is a largely tandem array of species specific repeats, in eukaryotes sequencing becomes difficult due to the size considerations; and (b) Repetitive DNA tends to be unstable in species like *Escherichia coli,* (when replicate bacterial artificial chromosomes) is attempted.

2.3 Centromere Aberrations: Disorganization of Cytogenetic Framework

When normal chromosomes are spliced and incorrectly repaired, then chromosomes with absence of centromere or addition of multiple centromeres could be formed. Such aberrant structures are undesirable because, they do not segregate properly and may often get lost from the dividing cell. In germ cells, this will result in unbalanced eggs or sperm (caused by extra or missing chromosomes or chromosome segments).

In humans and animals, errors in centromere may result in miscarriages, birth defects as well as, they may lead to cancerous cells. Centromere loss or extra centromeres would cause chromosomal segregation and may result in aneuploidy state (that is, the occurrence of one or a few chromosomes above or below the normal chromosome number). Rarely in humans, neocentromeres could be seen at new sites on a chromosome; and, currently over 60 known human neocentromeres have been identified on re-arranged marker chromosomes. This formation may be due to inactivation of centromere. The newly formed centromere is originally euchromatic (meaning lightly-packed form of chromatin as in DNA, RNA and protein) and lacks α-satellite DNA altogether. Any errors in centromere proteins may also result in auto immune diseases [2.8].

2.4 Chromatin – Structural and Functional Modifications: A Review

Another dimension of cellular complexity arises from chromatin constituents, the details of which are as follows: Human beings have 23 pairs of chromosomes per cell. That is, a total of 6 billion base pairs of DNA per cell. Normal human body consists of 50 trillion cells. Genetic material organization at molecular level is highly required by the cell for its normal functionality. In eukaryotic cell genetic material is well organized by compactly packing DNA with proteins called *histones,* in the nucleus hence, the achieved state is known as *chromatin.* DNA bounds to histones due to the existence of electrostatic forces between the negatively charged phosphate group in DNA and positively charged amino acids in histone proteins.

The eukaryotic cells when stained with basic dyes turned into bright color during cell division (mitosis and meiosis) due to its granular content i.e. condensed chromatin present in the nucleus. The term chromatin was coined by *Walther Flemming* in the year 1882 [2.9]. Chromatin is not found in prokaryotes (e.g., bacteria) that lack nucleus. Chromatin means "colored body". The fundamental unit of chromatin is Nucleosome, which is composed of DNA and histones. These nucleosomes, units of repeat ion are regularly spaced along a genome to form a nucleo-filament for higher level of compact organization and resulting into a chromosome. The main purpose of the chromatin existence is to regulate the biological processes such as DNA replication, gene expression, chromatin assembly and condensation, and cell division; the cell needs to change its structure of chromatin at certain specific regions of the genome and also coordinated time points, Most structural changes occur at nucleosome level, Chromatin histone modifications are very essential which are discussed in detail later.

Chromatin is the DNA of the nucleus and associated proteins such as histones. Mostly the protein consists of multiple copies of five types or kinds of histones. The amino acids arginine and lysine residues have free amino group that attracts the hydrogen ions giving tem positive charge. These amino acids tightly bind to the negatively charged phosphate groups of DNA. Chromatin might also consists of non-histone proteins such as transcriptional factors (TF) which are present in very small amounts, association of TF's with DNA is more transient. The five histone proteins variation from one cell type to another or from one species to another is very less when compared with the other non-histone proteins.

Histones keep the DNA organized, but also help to regulate expression of genes. Specifically modifications to histone proteins, such as methylation and acetylation help to regulate genes by activation or silencing. Later, chromatin code is read by the transcriptional regulators as histone modifications can modulate the accessibility of DNA to TF's. For example, β-globin gene regulation studies revealed that histone methylation might block DNA's access to TF's while acetylation might change the electrostatic interactions within the chromatin to open up DNA and allow gene transcription.

2.4.1 Structure of DNA and Chromatin

DNA molecule is a very flexible, based on the environmental conditions DNA can exist in many forms, there are three types of double helices DNA's in nature namely, A- DNA, B- DNA and Z –DNA. A and B forms of DNA are right handed forms whereas Z-DNA is left handed form. Hydrated DNA usually assumes B- form, whereas A-form is achieved when there is little water to interact with the helix and is also conformation adopted by the RNA. The Z-form has methylated deoxy-cytosine residues and occurs

during transcription where negative supercoiling stabilizes it. The most commonly occurred type of DNA in living organisms is B-DNA. DNA is deoxyribonucleic acid consisting of structural units called nucleotides, nucleotide consists of nitrogenous bases (Adenine-A and Thymine-T are pyrimidine bases while Cytosine-C and Guanine-G are purine bases and A), five carbon de-oxy ribose sugar and phosphate. The base pairing occurs in DNA, that is, GC pair is bound by three hydrogen bonds and AT pair is bound by two hydrogen bonds. Phosphate groups are joined by ester bond. In humans DNA is double stranded, the two strands run in opposite directions to each other therefore anti-parallel. The fundamental structural unit of chromatin is assemblage of DNA wound around the histone proteins.

Histones: These are present in eukaryotic cell nuclei, chemically they are highly alkaline and their function is to order the DNA into structural units called nucleosomes. Histone proteins have structural and functional role in transition of active and inactive chromatin states. There are two types of histones namely core histone and linker histones. Core histones consists of H1/H5, H2A, H2B, H3, and H4 and linker histones consists of H1 and H5. As indicated earlier, Chromatin consists of structural units so called nucleosomes, which consists of approximately 147 base pairs of DNA wrapped around a histone octamer consisting of two copies of each of the core histones H2A, H2B, H3, and H4, the linker histone H1 or H5 are usually positioned on top of the nucleosome for stabilizing higher order chromatin structure.

The four core histones, i.e. H2A, H2B, H3 and H4 are relatively similar in structure and are highly conserved through evolution. H2A and H2B form the dimers; and H3 and H4 constitute the tetramers. The helix turn helix motif of DNA and also the

feature of long tails on N- terminal end on aminoacid structure is common feature in core histone. Histones have maintained higher degree of conservation in the course of evolution; however the histone variants in the nucleosomal octameric core that had evolved played diverse roles in the gene regulation and epigenetic silencing.

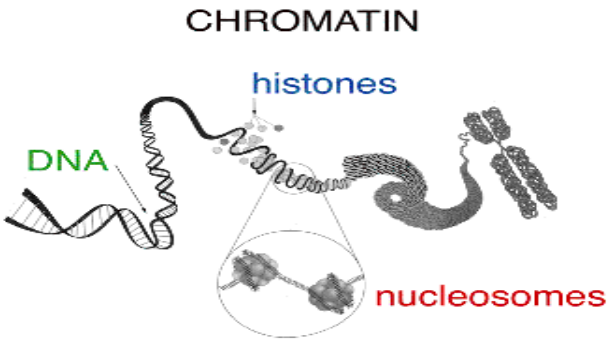

Figure: 2.3 Structure of chromatin

(Adapted from faculty.jsd.claremont.edu, Available at: http://faculty.jsd.claremont.edu/jarmstrong/researchint.htm)

Histones binding to DNA critically depend on the amino acid sequence of the histone and do not depend on particular nucleotide sequences in the DNA. Histones are highly conserved molecules during the course of evolution. For example, Histone H4 in the calf (young cow) differs from pea plant at only two amino acid residues in the chain 102.

Although amino acid sequence is same, each histone molecule differs in structure due to its chemical modifications that occur later to individual amino acids, for example acetyl groups to lysines, phosphate groups to serines and threonines and methyl groups to lysines and arginines. 75 to 80% of histone molecule is present in its core, the remaining

percentage at the N-terminal dangles out from the core as histone tail. The chemical modifications occur on these tails especially of H3 and H4 and these changes are reversible. Histone modifications affect chromosome function through at least two distinct mechanisms. Primarily, histone modifications may alter the electrostatic charge of the histone resulting in a structural change in histones or their binding to DNA and secondly, modifications are binding sites for protein recognition modules, such as the bromodomains or chromodomains that recognize acetylated lysines or methylated lysine, respectively. Histones are subjected to a wide variety of posttranslational modifications including but not limited to lysine acetylation, lysine and arginine methylation, serine and threonine phosphorylation, and lysine ubiquitination and sumoylation, these modifications occur within the histone at N-terminal tails protruding from the surface of the nucleosome.

Nucleosome: It is a subunit of chromatin composed of short length of DNA wrapped around a core of histone proteins. The human genome contains 23 chromosomes i.e. approximately 3 billion nucleotide pairs therefore compact organization is very important each nucleosome is about 11nm in diameter , the DNA double helix wraps around a central core of eight histone proteins to form a single nucleosome, the second histone protein (H1) fastens the DNA to nucleosome core explained below in detail [2.9].

Chromatin is the repeating units of nucleosomes, which consist of ~147 base pairs of DNA wrapped around a histone octamer consisting of two copies of each of the core histones H2A, H2B, H3, and H4. Linker histone H1 is positioned on top of the nucleosome core particles stabilizing higher order chromatin structure. The changes in chromatin structure are effected by modifications that are very predominant in the core

histones at their N- terminal tails. The chromatin modifications are due to acetylation, methylation, phosphorylation, ubiquitylation, sumoylation, ribosylation, deamination and isomerization. Enzymes play a key role in catalyzing these chemical reactions. For example the chemical compounds like methyl or acetyl groups can be covalently linked to certain amino acid residues, such as lysine or arginine, further resulting in the modification of amino acids serve as docking sites (active site) for other proteins, which specifically recognize the histone mark and enable them to wrap or unwrap during the cell cycle such as mitosis, meiosis and interphases [2.10]. The total mass of a single nucleosome complex is about 100,000 daltons. In eukaryotes the higher order of packing gives the chromosome its compact structure i.e. typical human chromosome is about 0.1cm in length and would span the nucleus 100 times.

2.4.2 Chromatin Organization

Based on microscopic observations, two-levels of chromatin organization are indicated: They are: (i) *Heterochromatin and (ii) Euchromatin*. They refer to states of compaction (DNA and histones) and their transcriptional potential. Heterochromatin is tightly coiled form of DNA mean's condensed, genetically inactive but may play a role in controlling metabolic activities, transcription and cell division (Interphase). Chromatin in this state stains darkly in karyograms (Heitz, 1928) [2.11]. Heterochromatin is usually localized on periphery of the nucleus in a eukaryotic cell. Heterochromatin does not alter its structure in condensation throughout the cell cycle. There are two types of heterochromatin, namely *constitutive* and *facultative heterochromatin* depending on the DNA that they contain.

Constitutive heterochromatin predominantly occurs at centromere and telomeres. They usually exist as highly condensed state with repetitive DNA and are largely transcriptionally silent. As constitutive heterochromatin is genetically inactive mostly, thus serves as a structural element of the chromosome.

Facultative heterochromatin is a non-repetitive and inactive DNA, but under specific developmental or environmental signals, loses its condensed structure and become transcriptionally active, for example the inactive X chromosome in female somatic cells.

In general, heterochromatin is greatly enriched with transposons and other junk DNA and usually replicates late in S-phase of the cell cycle. Genes present in heterochromatin are generally inactive (no transcription) and increased methylation of cytosines in CpG islands of the gene's promoter is observed.

Histones in the nucleosomes of the heterochromatin shows decreased acetylation and increased methylation of lysine-9 in histone H3 (H3K9) providing binding site for heterochromatin protein (HP1) which in turn blocks transcriptional factors need for gene transcription. Increased methylation of lysine-27 in histone 27 in histone H3 (H3k27) is also observed.

Euchromatin: In euchromatin DNA and histones are loosely or lightly packed, mostly transcriptionally active. Chromatin in this state stains lightly in karyograms (a diagram of chromosomes of a cell arranged in homologous pairs). It is found in both eukaryotic and prokaryotic cells (even wit out nuclei), but heterochromatin is found in eukaryotes. The cells can transform euchromatin into heterochromatin and vice versa, the reason behind this transformation is to control gene expression and replication, as these

processes behave differently in compact and condenses states of chromatin (*accessibility hypothesis*). This transformation of chromatin state is due to the chemical reactions occurring on the N-terminal of the histone tails by the action of specific enzymes. In general euchromatin is loosely packed in loops of 30 nm fibers, genes are active and decreased methylation of cytosines in CpG islands of the gene promoters and lysine-9 and lysine-27 in histone H3. Heterochromatin and euchromatin differ in their biophysical conformations and in metabolic expression of their genes but same in their basic structure of DNA arranged within chromosomes [2.11].

*Pre-nucleosomes***:** A novel chromatin particle was discovered recently by biologist *James Kadonaga* [2.12] named pre-nucleosomes, which is a precursor of nucleosome made up of intermediate DNA – histone complex. The nucleosome is the basic repeating unit of chromatin. The pre-nucleosomes are converted into nucleosomes by motor proteins that use the energy molecule ATP. The packing of DNA with histone proteins to form chromatin plays key role in stabilizing chromosomes and regulation of genes mainly in DNA replication. The pre-nucleosome is likely to be an important player in how our genetic material is duplicated and used [2.12].

2.4.3 Abnormal Features of Chromatin

Epigenetics is the study of heritable changes in the chromatin without involving the changes in the DNA sequences. The chromatin histone proteins associated with DNA may be activated or silenced. In a multicellular organism, the differentiated cells express only genes that are necessary for their own activity. Epigenetic changes in an organism occurs in its lifetime, but if any mutations occur in the DNA of the sperm or egg that undergoes fertilization, then some epigenetic changes are inherited from one generation

to the next [2.13]. The aberrant chromatin state is the cause of disease; hence by unrevealing the chromatin structure and its functionality it is easy for us to understand many diseases including developmental disorders and tumors. Chromatin plays key role in the all aspects of cell behavior (transcription, translation and protein synthesis). Embryonic development and stem cell behavior is regulated by chromatin. It also impacts the cell cycle including chromatin condensation during mitosis and DNA replication during S phase. Chromatin studies in stem cells such as embryonic stem cells help us to better understand the induced pluripotent stem cells. As stem cells and tumor cells have unique chromatin structure, more active state of chromatin compared to normal cells. Breakthroughs can be achieved in the fields such as Cancer biology, regenerative medicine etc. [2.14].

2.5 Closure

This chapter is written as an overview on cytogenetic complex based on a number of archival literature details. Its contents are useful toward understanding the studies presented in the ensuing chapters.

CHAPTER III

CHROMOSOMAL ABERRATIONS

3.1 Introduction

As indicated in earlier chapters, the present study is a cytogenetic-inspired approach to determine quantitatively the extent of the presence of chromosomal aberrations. In this context, it is necessary to identify first existing measures that are traditionally used to quantify the *chromosomal abnormality* (CA). The following sections outline the details available thereof in the archival literature.

3.2 Chromosomal Aberrations: An overview

There are two regimes of quantifying chromosome abnormality. They correspond to: (i) At nucleotide level and (ii) at cytogenetic level**.** Relevant details are as follows:

CA at nucleotide level: The variations account approximately 12% of human genomic DNA and the genetic variations such as deletion, duplication or inversion, these might range from one kilo base to several mega bases in size whereas single nucleotide polymorphisms (SNPs) differ from what is known as *Copy number variations* (CNVs) as the effect is only on one single nucleotide base. CNVs are mutations that might include deletion, duplication or inversion.

Human genome consists of nearly 6 billion nucleotides of DNA packaged into two sets of 23 chromosomes. One set inherited from each parent. The segments of DNA ranging in size from thousands to millions of DNA bases might vary in copy number.

CNVs might even lead genes to dosage imbalances. CNVs sometimes might not be directly related to cause of disease in particular. In some cases CNVs may influence gene expression, phenotypic variation and adoption by altering the gene dosage. Hence CNVs play an important role in human disease as well as drug response. Gene dosage describes the number of copies of a gene in a cell and gene expression can be influenced by higher and lower dosages, for example deletion lowers the gene dosage. Immune system and brain development are mostly enriched by CNVs in the course of evolution.

CA at cytogenetic level: CNVs mentioned earlier depict a kind of structural variations having an abnormal number of copies of one or more sections of the DNA in chromosomes of the cells, CNVs includes large regions of genome that ate involved in insertions, deletion, duplications and translocations. Unequal recombination's also leads to CNVs. this variation is about 12% of human genomic DNA and each variation ranges from one kilobases (kb) to several megabases (mb) in size. CNVs are caused by inheritance or *de nova* mutation (i.e. genetic mutation that neither parent possessed nor transmitted). CNVs were first unrevealed by Human genome project, that estimated 0.4% of the genomes of unrelated people typically differ with respect to copy number. De novo mutations have been observed in identical twins even though they have identical genomes. CNVs have been associated with susceptibility or resistance to disease [3.1].

CNVs play a role in evolutionary adaption in humans as well as other mammals. For example human salivary amylase gene (AMY1) is present in 6 to 15 copies, which plays a key role in adoption to high starch diet that improves the ability to digest starchy food, whereas chimpanzees only have two diploid copies of AMY gene [3.2]. There are

two types of CNVs Broad events consisting of several Mb or even the whole chromosome while focused events, consisting of normally restricted to few Mb.

CNVs can be detected by various types of tests: fluorescence in situ hybridization, comparative genomic hybridization, high-resolution array-based tests based on array comparative genomic hybridization, and quantitative-PCR based technique for analysis and/or validation of known CNVs. The most efficient method used in detecting CNVs is array-based method or Virtual Karyotype. BAC (Bacterial Artificial Chromosome) arrays are historically the first micro array methods used in detecting DNA copy number analysis.

Single nucleotide polymorphisms (SNPs): SNP is a DNA sequence variation that occurs at a single nucleotide. Genetic code is specified by four nucleotides namely, adenine, thymine, cytosine, and guanine (A, T, C, G) in the genome. SNPs occur mostly in non-coding regions compared to coding sequence. SNPs occur normally throughout the individual DNA. SNPs occur in every 300 nucleotides on average; approximately 10 million SNPs are present in the human genome. The SNPs are found in DNA between genes. The SNPs that occur within a gene or in a regulatory region near a gene play direct role in diseases by affecting gene function.

Mostly SNPs do not have effect on human health but it might be helpful to understand drug response, susceptibility to environmental factors such as toxins and risk of developing particular diseases. SNPs found in coding region are particularly of interest as they alter the biological function of a protein, only 3-5 % of the human DNA sequence codes for the production of protein while the rest remains non coding region where the occurrence of SNPs might not be much of interest.

SNPs are genetically stable and occur frequently throughout the genome. They act as “biological markers”, meaning that these DNA segments are present within an identifiable physical location that can be easily located and used to construct chromosome map showing the positions of known genes or markers. 99.9% of one individual DNA sequence is identical to another individual, only 0.1% difference is present, of which 80% are contributed by SNPs. Identification of genetic components of disease is facilitated by studying SNPs.

SNPs occur when DNA replication enzymes make an error as they copy the cell’s DNA during meiosis; hence the enzyme incorporates approximately one mistake in 9-10 million nucleotide bases. Some SNPs that are harmful cause diabetics, cancer, heart disease, Huntington’s disease and hemophilia. Sometimes changes in each gene become apparent under certain conditions leading to susceptibility of lung cancer.

Techniques to detect SNPs are hybridization techniques that include micro arrays and real time PCR. Enzyme based techniques such as nucleotide extension, cleavage, ligation and direct sequencing are few techniques used to detect SNPs present in the human genome (a genome represents all the genetic material within the chromosomes and the so-called transcriptome depicts the entire set of gene transcripts. Following the central dogma hierarchy, proteome denotes the entire set of proteins) [3.3].

3.3 Features of Chromosomal Aberrations

Chromosomal abnormalities as discussed earlier are result from either a variation in the chromosome number or from structural changes. These changes might have occurred spontaneously or induce by environmental agents such as chemicals, radiation etc., Mutations occur when there is mistakes when genes are copied as cell division occurs to

produce new cells. Chromosomal abnormalities may involve autosomes, sex chromosomes, or both.

Cytogenetics, is used to understand chromosomal disorder and their relationship to health and disease. Cytogenetic analysis are those diagnostic methods to analyze these disorders such as prenatal diagnosis, multiple birth defects, and abnormal sex development and in some cases of infertility or multiple miscarriages, cancer, hematological disorders. The types of chromosomal disorders that can be detected by cytogenetics are numerical aberrations, translocations, duplications, deletions, and inversions.

Identification of individual chromosomes was difficult until the discovery of staining techniques such as Q-banding reveals the structural organization of chromosomes. The banding patterns are important in recognizing the genetic disorder. Chromosomal abnormalities of even a minute segment or band are now known to be the basis for number of genetic diseases.

Rearrangement: Chromosomal rearrangements can cause cancer by forming a hybrid gene or by disregulation of gene. (Hybrid genes are fusion of two different genes at the rearrangement break points). The aberrant chromosome formed due to rearrangement is a hybrid gene, which in turn codes for aberrant protein that disrupts the normal functionality of the cytogenetic system, for example, chronic myeloid leukemia.

Fluorescence in situ hybridization (FISH) is the most common method used to detect chromosomal rearrangements. More than 200 different types of fusions have been identified, they often involve oncogenes that cause cancer (MLL, RET and EWSR1 genes).hybrid genes are not specific to particular type of cancer, as the same hybrid gene involved in multiple types of cancer, hence these genes might initiate cancer progression

in a variety of different tissues. Disregulation of normal genes can cause the conversion of normal cells into cancerous cells. Overexpression of gene is usually observed in the disregulation. Mutations in the human DNA repair genes can deregulate normal cellular processes as well as active gene when transposed might result in activation of silent gene [3.4].

Clonal or non-clonal chromosomes: A part of cytogenetic characterization and classification of chromosomal aberration involves elucidating normal karyotypes and identifying major recurrent chromosomal aberrations. Cells that accumulate mutations correspond to a somatic evolution and implicate natural processes like aging and development of diseases like cancer. The survival of the cell in the somatic evolution with acquired mutation is decided by the increased fitness of the cells.

Cells in neoplasmic (tumorous condition) try to increase their fitness with the resources like oxygen, glucose and space, further they generate more daughter cells competitively as compared to cells that lack mutation, considering such mutant cells called (clone) with the available resources expand in the neoplasm. Relevant clonal expansions form the signature of natural selection in cancer and correspondingly the population of aberrated cells (that promote the neoplasmic) cancerous conditions are clonal chromosomes.

The biological significance of clonal karyotypic abnormalities in neoplasms indicated above has been understood to a fair extent. However there are chromosome aberration that are not consistent within the cytogenetic definition of a clone, but do exist as an artifact manifesting in random losses of cells and causing karyotypic

instability/evolution such, non-clonal abnormalities may predict future, clinically significant clones being ignored as discussed in [3.5].

Contamination karyotypes: In cytogenetic context molecular karyotyping involving tissue culture often shown the presence of contaminations that limit the efficacy of karyotype evaluations. For example, details pertinent to karyotyping products of conception with maternal cell contamination. Relevant studies are useful for fetal welfare. Thus, contamination karyotypes implies cross contamination of cytogenetic entities of two systems such as maternal and fetal.

3.4 Chromosomal Aberrations and Genetic Disorders

Genetic disorder is a diseased state caused by abnormalities in genes or chromosomes. Usually they are present from before birth. Sometimes genetic disorders are passed down from parent's genes or might be caused by new mutation changes to the DNA. For example same type of cancer can be caused by inherited genetic condition or by mutation due to non-genetic causes.

Single Gene disorder: nearly 4000 human diseases are caused by single gene disorder. The cause of single gene disorder is due to mutated gene. Based on chromosomal location the genetic disorders can be classified into autosomal and X-linked types. Further these are sub-divided into Autosomal dominant, Autosomal recessive, X-linked dominant and X-linked recessive based on their allelic expression.

3.5 Chromosomal Aberrations: Oncological Considerations

Accumulation of genetic alterations in the cells causes human cancer. Chromosomal abnormality associated with cancer was first discovered by Peter Nowell

and David Hungerford in 1960 [3.6] in chronic myeloid leukemia (unrestricted growth of myeloid cells in the bone marrow).

In 1969, Foulds [3.7] had studied about cancer and its stage evolution, and later in 1982, Berenblum [3.7] established three distinct stages: the initiation stage, the promotion stage and the progression stage**.** If the first two stages underlie the triggering of cell transformation, the third stage determines the transformation of a benign tumor into a malignant form, with the maintenance and evolution of malignancy.

Proto-oncogenes encode proteins that are involved in the control of cell growth. Alteration of the structure and/or expression of proto-oncogenes can activate them to become oncogenes capable of inducing in susceptible cells the neoplastic phenotype (neoplasm means abnormal proliferation of cells, while Neoplasms may be benign, pre-malignant (carcinoma in situ) or malignant (cancer). Oncogenes can be classified into five groups based on the functional and biochemical properties of protein products and their normal counterparts (proto-oncogenes). These groups include growth-factors, growth-factor receptors, signal transducers, transcription-factors, and others such as programmed cell-death regulators [3.8].

Models of genetic abnormality representing the cancer: Cancer is caused by a sequence of genetic abnormalities arising in a tumor cell. Understanding the order relevant of occurrences for the staging of tumors is very crucial in cancer treatment. The genetic changes influence tumor progression either by deactivation of tumor suppressor genes (that increases the probability of further genetic changes) or by activation of oncogenes (that gives the cell cancerous properties). The model depicting occurrence of genetic

abnormalities can be divided into two types namely, the linear structure model and oncogenetic tree-based structure model.

Linear model: The genetic profile of individual tumors varies widely because no single mutation is present in all tumors; certain genetic changes tend to occur early in the development and others relatively late. The steps involved in carcinogenesis were first studied by Vogelstein et al. [3.9] in colorectal tumorgenesis as a preferred order of occurrence of the genetic abnormalities while acknowledging the existence of other pathways.

Oncogenetic tree model: The so-called tree model represents a combination of several pathways in single model meaning that certain genetic abnormalities may lead to several other changes due to increased chance of occurrence; whereas, in the linear model, it is quite opposite. Hence, oncogenetic tress allows multiple possible pathways and parallel progression along several pathways in the same tumor. Oncogenetic trees include linear model as a special case. Oncogenetic tree model was first studied by Desper et al. [3.10]. It is subdivided into branching trees and distance based trees. Cancer is genetically heterogeneous even in tumors that are considered to be clinically or pathologically homogeneous. For example, renal cell carcinoma has a high degree of genetic heterogeneity. Hence, tree models capture heterogeneity. Moreover they are not dependent on a single linear progression, so tree models are more flexible and realistic when compared to a linear model.

Branching trees: In a branching tree, there is one node called a root, and every other node denotes one of the events. An edge depicted as (i : j) represents a hypothetical cause-and-effect relationship meaning that the occurrence of event i makes the occurrence of j more

likely. Thus, heterogeneous possibilities for how oncogenesis can progress can be represented. The choice of which edge to include in the branching tree is based on a weight function that takes into account how often each event occurs and how often each pair of events take place together in the same tumor.

Distance-based trees: The distance-based trees have all of the events at the leaves, whereas the internal nodes are hidden, unnamed events (much like a phylogenetic tree has the existing species) as leaves, and the hypothetical common ancestors as internal nodes. Another significant difference is that, in such trees, each edge has a length, and the trees are drawn so that the horizontal distance along an edge is proportional to length. The trees are constructed by first defining an (n x n) distance matrix that describes for each pair of events whether they tend to occur together or not. The second step is to use existing phylogenetic methods to find the phylogenetic tree that best fits the distance matrix. For this reason the trees constructed by this method are called "distance- based trees."

The event labels A, B, C, D, E, and F in Figure 3.1 represent CGH aberrations. The CGH or *comparative genomic hybridization* (also known as *chromosomal microarray analysis* CMA) is a molecular-cytogenetic method for the analysis of copy number changes (gains/losses) in the DNA content of a given subject's DNA and often in tumor cells. CGH will detect only unbalanced chromosomal changes. Structural chromosome aberrations such as balanced reciprocal translocations or inversions cannot be detected, as they do not change the copy number.

In Figure 3.1, cancer progression is from left to right. The top two models are examples of branching models because all vertices have a label. The top model is a path

and a tree; the middle model denotes only a tree, but not a path and the bottom model does not represent a tree because there are two paths stem from the root to event E.

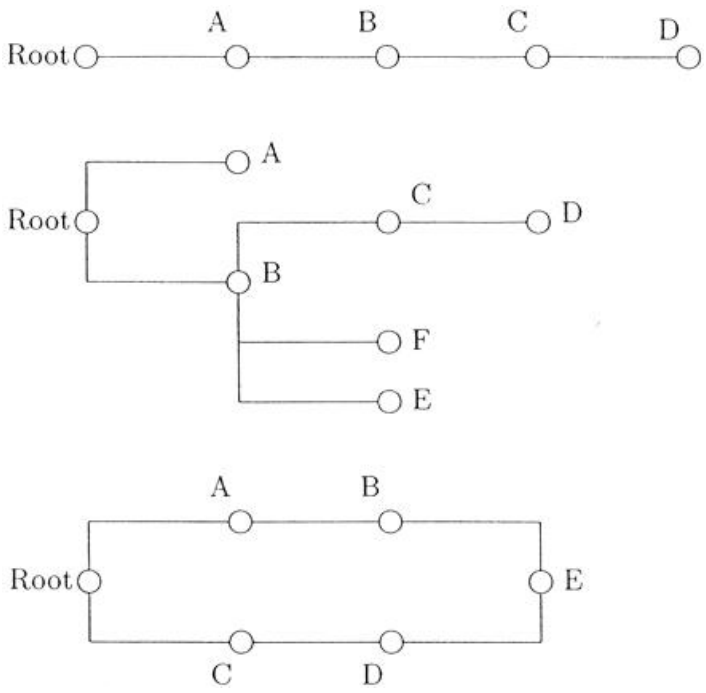

Figure: 3.1 Oncogenic Tree Models [3.10]

3.6 Closure

Commensurate with the objectives of the present study, this chapter offers an outline featuring details pertinent to chromosomal aberrations. Pertinent information includes definitions on the types of CA, their implications on genetic disorders and oncological states. A gamut of citations is presented on the topics addressed.

CHAPTER IV

MIXTURE OF NORMAL AND ABNORMAL CHROMOSOMES: EVALUATION OF CHROMOSOMAL ABNORMALITY VIA SIMPLE PROPORTION MIXTURE MODEL

4.1 Introduction

The traditional approach in cytogenetics that measures cell contamination is based on representing the cell-contents by simple fractions of normal and abnormal chromosomes being present. Relevantly, reviewed in this chapter is such a traditional proportion fraction concept applied to the mixed states of normal and abnormal chromosomes in a cytogenetic complex and evaluate the global pathogenic state of genetic disorders. For example, indicated in [4.1], is an estimation of the incidence of chromosomal abnormalities (for a given male/female sex ratio of conception) among spontaneously observed abortions masked by maternal cell contamination (MCC). The results obtained in [4.1] is reproduced in Table 4.1. The analysis pursued in [4.1] leads to determining the frequency (or relative percentage) of spontaneous abortion *versus* MCC under the conditions such as those indicated in Table 4.1.

Essentially, the traditional approach as in [4.1] of analyzing a mixture of normal cellular contents with contaminations such as abnormal chromosomes relies on specifying each content in terms of their relative percentage that is, in terms of simple prorated values. Suppose the normal chromosomes content is N_c % and the percentage of aberrated chromosome is N_{ac}% so that the total ($N_c + N_{ac}$) = 100%. Then, the data acquired clinically

or otherwise denote the underlying inferential statistics to depicting the associated variances on N_c and $N_{ac.}$

Table 4.1: Possible maternal cell contamination details in two groups of "46,XX" spontaneous abortions with or without Y chromosome [4.1]

Exogenous characteristics	Symbol	"Y +" spontaneous abortions (n = 18) With Y chromosome	"Y -" spontaneous abortions (n = 94) Without Y chromosome
Gestation	X_1	8.88 ± 2.60	8.86 ± 2.76
Maternal age	X_2	26.00 ± 6.54	25.54 ± 5.27
Paternal age	X_3	27.21 ± 6.36	28.28 ± 5.58
Diagnosis: Blighted ovum	X_4	4	20
Diagnosis: Missed abortion	X_5	12	62
Diagnosis: Others	X_6	2	12
Tissue type for culture initiation: Extra-embryonic mesoderm	X_7	14	64
Tissue type for culture initiation: Chorionic villi	X_8	3	30
Tissue type for culture initiation: Duration of long-term culture	X_9	35.08 ± 16.88	24.05 ± 13.86

The study due to [4.1], the data set as in Table 4.1 supposedly accounts for its full range of statistical variations on the class of spontaneous abortions observed. However, such statistical inferences [4.1] are limited in presenting the associated statistical error-bar on the observed details. That is, the observed inferences in [4.1] are confined to, almost a deterministic specification. But, in reality, any statistical inference should be specified within a stretch of error-bar on the observed variable *versus* the input variables. Hence indicated below is a method to accommodate error-bar specification on the analysis such as in [4.1].

Given a data set as in Table 4.1, the net effect of the proportional content of abnormal chromosomes can be determined by logistic regression of the details on various exogenous entities involved. That is, relevant data can be logit-regressed in order to find the net risk-factor due to all the exogenous characteristics listed as X_1 to X_9 in Table 4.1. All such factors will then be collectively accounted for in determining the cumulative MCC based risk-factor on observed abortions. The outcomes will correspond to an error-based stretch of details. The method of logistic-regression is outlined below:

Denoting $z = (X_1 + X_2 \ldots + X_9)$, the aforesaid risk-factor can be defined in terms of logistic-regression function, $f(z) = [1/1 + \exp(-z)]$ by randomly changing the ensemble value of $\{X_i\}_{1\ldots.9.}$

Notwithstanding, the classical representation of $f(z) = [1/1 + \exp(-z)]$, when the variable set of exogenous factors constituting z is statistically random, f(z) can be written modified indicated by Dupont in [4.2]: The modified logit function $f(z) = (1/2) + (1/2)L_q(z/2)$ with q being a stochastical order-function denoting the random feature of the

statistics involved. When $q \to 1/2$, it corresponds to a total disorder and when $q \to \infty$, the system depicts an ordered entity. Accordingly, a pair of risk-factor variations *versus* an ensemble of random-trials on the set $\{X_i\}$ (concerning MCC-specific characteristics) can be obtained to represent the upper and lower bounds of risk statistics involved. Relevant details on the computations are presented below.

4.2 Cytogenetic Mixture Contents: Statistical Implications

The considerations on log-regressing the data as applied to a cytogenetic complex are illustrated in this study with an example pertinent to the details of [4.1]. That is, the problem of assessing the net contamination details *via* logistic function regressed risk-factor is indicated here with reference to clinical data on maternal cell contamination described in [4.1].

4.3 A Case Study:

Suppose a model is considered depicting the cytogenetic complex with its constituents being regarded as a mixture of simple proportion either by volume or by weight. Hence considered in [4.1] is a heuristic approach in which a cell is modeled with items of contaminated and uncontaminated nature within certain proportion. Hence the net effect of contamination on the overall abnormal feature of the cytogenetic complex is evaluated. Specifically considering a maternal cell contamination depicting certain undesired cellular entity (leading to spontaneous abortions) is considered and a simple proportionality model is indicated on the incidence of MCC *versus* chromosomal abnormality masking the possibilities of spontaneous abortions. The model in [4.1] eventually provides an estimation of the risk involved as a result of MCC. The data used thereof corresponds to clinical samples from 97 patients with aborted embryos.

The risk factor deduced as above, however conforms to a single value deduced from the clinical data. It ignores various associative considerations of statistical nature such as gestational age, maternal and paternal ages, diagnostic details on blighted ovum and missed abortion etc., tissue type used for culture initiation and duration of long term culture. All these characteristics are mostly random but within specified bounds and they constitute the exogenic variable $\{X_i\}$ indicated earlier in making of a log regression function f (z) with $z = \Sigma_i X_i$.

In order to accommodate the statistical notion to the underlying features in designing f (z), the method due to [4.2] is invoked to obtain the risk factor evaluation with its error-bounds specified by upper and lower limits.

4.3.1 Case Study Example: Maternal Cell Contamination [4.1]

This problem is concerned with developing a mathematical model to evaluate the MCC, which causes spontaneous abortions as discussed in [4.1]. Hence, the significance of cytogenetic analysis of prenatal selection factors on fetus health is studied.

Presented in [4.1] is an analytical model that estimates the incidence of chromosomal abnormalities specific to MCC. A high rate of MCC may distort the state of chromosomal abnormality encountered in spontaneous abortions across the first trimester of pregnancy. In [4.1], a method is indicated to estimate such MCC effects of cytogenetic interest in prenatal situations *vis-à-vis* abortion implications.

In the MCC model due to [4.1], N denotes the number of samples of spontaneous abortions observed with four major versions of chromosomal constituents namely, (A: 46, XX), (B: female spontaneous abortions with chromosomal abnormalities), (C: 46, XY), and (D: male spontaneous abortions with chromosomal abnormalities).

Hence, N = (A + B + C + D), where the set A, B, C, D are karyotypes with A being the chromosomes in the maternal cells. The chromosome A: 46, XX is further subdivided into four types as follows with reference to the sex-ratio of the embryo: A_{fn} : 46, XX (female normal); A_{fa} : female spontaneous abortions with chromosomal abnormalities (female abnormal); A_{mn} :46, XY (male normal); A_{ma} : male spontaneous abortions with chromosomal abnormalities (male abnormal); As such, it follows that:

$$A = (A_{fn} + A_{fa} + A_{mn} + A_{ma}) \tag{4.1}$$

For a representative sample of N, the relative proportions of B, C and D and A_{fn} and A_{fa} are indicated in [4.1]; and, the results are reproduced in Table 4.2

4.3.2 Computational Details and Discussion on Case Study Example

The mathematical model in [4.1] is based on the structure of cytogenetic factors expressed in terms of proportions of various chromosomal entities identified earlier as A, B, C and D; hence, the factor for MCC is indicated as k = (C+D)/(C+D+B). It defines the probability of male embryo detection in the A group. The influence of MCC as a function of k is then estimated on the spontaneous abortions observed.

For a specific set of details presented in Table 4.2, and considering the karyotypes 46, XX relevant representative values on B, C and D are as follows: B (abnormal female = 139), D (abnormal male = 94) and C (46, XY = 86). These are typical sample sizes of the observed numbers in the model of MCC presented in [4.1]. Hence the corresponding factor of MCC, namely k = 0.564. Relevant to this value of k, the risk-factor resulting from the influence of MCC is shown in [4.1], by a single value approximately equal to 0.95% over a range of k between 0.2 to 0.5, and, the associated risk of spontaneous abortion corresponds to, 50% - 95%.

Table 4.2 Proportional Modeling the structure of cytogenetic factors of prenatal selection corrected for cell contamination [4.1]

Cytogenetic factor	Observed value	Expected value
46, XX frequency	A/N	[A(1 - *k*) - (A B *k*)/(C+D)]/N
46, XY frequency	C/N	[C{1+ (A *k*)/(C+D)}]/N
Frequency of female spontaneous abortions with chromosomal abnormalities	B/N	[B{1+ (A *k*)/(C+D)}]/N
Frequency of male spontaneous abortions with chromosomal abnormalities	D/N	[D{1+ (A *k*)/(C+D)}]/N
Frequency of chromosomal abnormalities in total sample	(B+D)/N	[(B+D){1+ (A *k*)/(C+D)}]/N
Frequency of chromosomal abnormalities in the "46 XX" group	-	*k*(B+D)/(C+D)
Sex ratio in spontaneous abortions with normal karyotype	C/A	[C{1+ (A *k*)/(C+D)}]/ [A(1-*k*)-(A B *k*)/(C+D)]

The simple proportion-based analysis due to [4.1] gives only a rigid span of results on the risk-factor as above. In contrast, suppose the risk-factor is deduced on the basis of f (z), the results can be specified error-bounded with upper and lower levels as illustrated in

Figure 4.1. Here, in order to specify the error-bounds, f(z) is taken as follows: f(z) = ½ + ½ Lq (z/2) as discussed in [4.2], where q → ½ denotes the upper-bound and q → ∞ depicts the lower-bound. Further, Lq (x) denotes the so-called Langevin-Bernoulli function given by: Lq (x) = (1+1/2) coth [(1+1/q) x] – (1/q) coth [(1/q) x].

4.4 Results

4.4.1 Case-1a: "Y+" Spontaneous Abortions

Presented in Table 4.1 are details concerning possible MCC as regard to two groups of "46, XX" spontaneous abortions with and without Y chromosomes indicated respectively as (Y +) and (Y −).

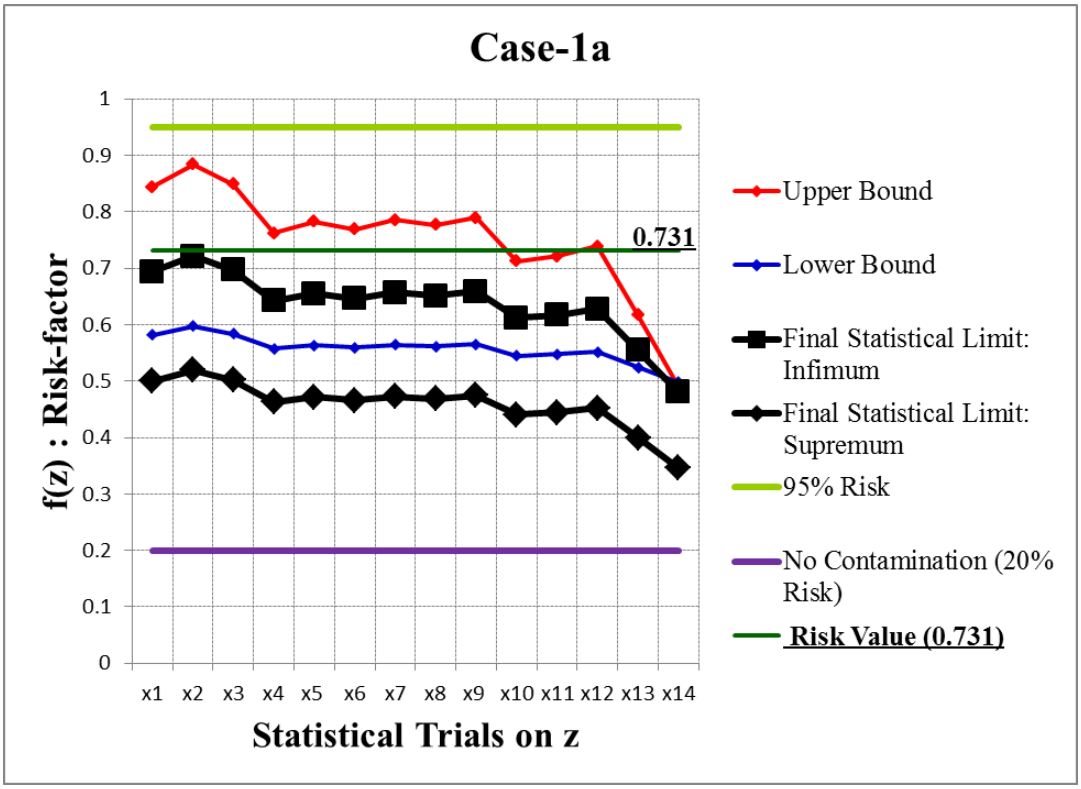

Figure 4.1 (a) A risk-factor versus an ensemble of trials (T_1, T_2……T_{14}) of data presented in Table 4.1 for (Y+) state

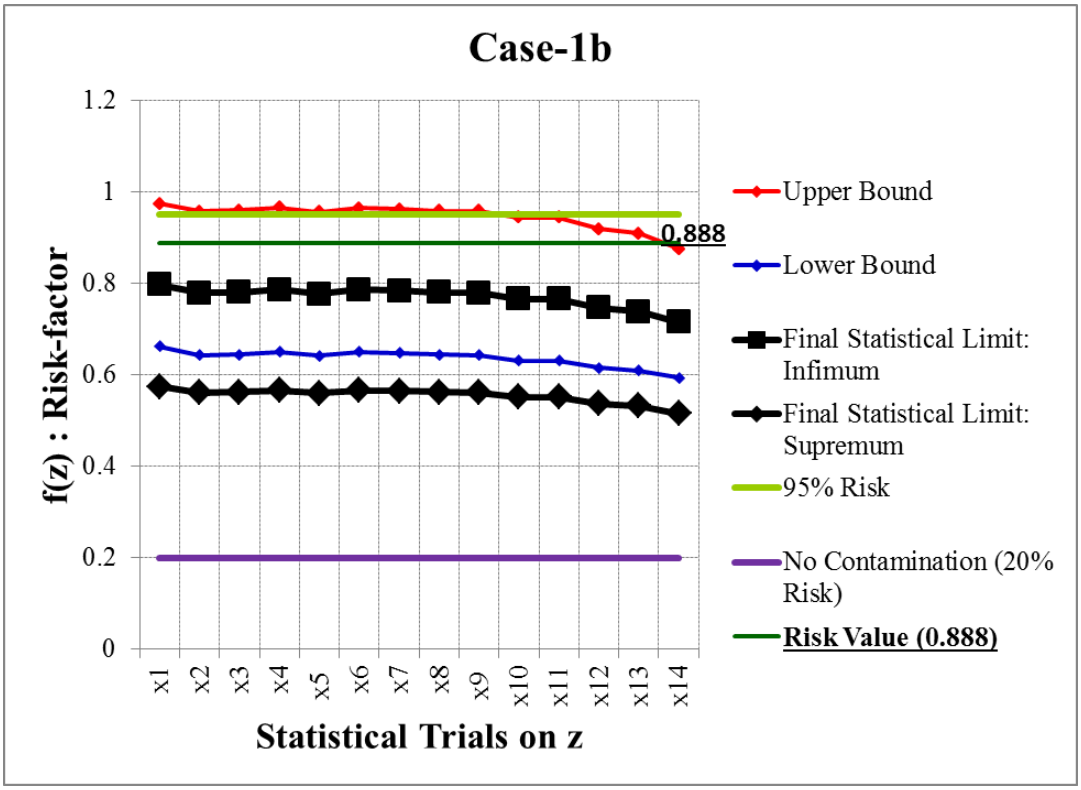

Figure 4.1 (b): Risk-factor *versus* an ensemble of trials (T_1, T_2……T_{14}) of data presented in Table 4.1 for (Y –) case.

With reference to Table 4.1 and (Y+) status, the sum total of exogenous characteristics (that decide the underlying risk) namely, $z = \sum_i X_{i=1,2,...9}$ is determined for a nominal set of $\{X_i\}$. Hence, f(z) = ½ + ½ L_q (z/2) is evaluated. Inasmuch as Table 4.1 suggests the existence of a span of deviation with respect to the nominal values of each X_i, a number of random trails on each X_i can be specified; and corresponding f(z) is evaluated for each ensemble set. (For example, with X_1 = 8.88 taken as a nominal value, it is varied randomly in each trial over the deviation, ± 2.60 indicated in Table 4.1). Hence, the evaluated risk-factor deduced *via* f(z) = ½ + ½ L_q (z/2) over an ensemble of several

statistical trials is plotted in Figure 4.1 (a); further, for q = ½ (upper-bound) and q → ∞ (lower bound) corresponding computed data are also shown in Figure 4.1 (A).

In addition, the *infimum* and *supremum* pair is deduced on the bounds assuming a statistical quantile of 48 % about the mean value.

4.4.2 Case-1b:"Y−" Spontaneous Abortions

Again, with reference to (Y −) case indicated in Table 4.1, similar computations done for (Y +) are repeated and the results are illustrated in Figure 4.1(B)

4.5 Discussion and Closure

Relevant to the results obtained and presented the following observations can be made: (i) with no MCC being present the associated risk is as low as 20% and the highest risk is specified as 95%. (ii) in reality taking the statistical variation of the various factors (X_i) of the contaminated state allows the prediction of the risk *via* logistic regression. Correspondingly an upper and lower bound of the risk-factor involved is deduced along with the associated infimum and supremum limits. As illustrated in figures 4.1 (a) and 4.1 (b). In the studies due to [4.1] the model indicated signifies the extent of possible spontaneous abortion resulting from MCC under various prenatal selection factors. For a given set of A, B, C, D and N, MCC (k) and male, female ratio as indicated in table 4.1, the associated risk value for Y+ case evaluated by logistic regression is 0.731 and corresponding value for Y− is 0.888. In contrast the present study rather specifies an error range which can be considered as more realistic.

Thus the present study is based on proportional mixture considerations on the contaminated and uncontaminated entities in cultured cells. It provides thereof the range of risk involved (specified within an error bound on the possible contamination

related risk-factor of possible abortion). The analysis and the computation of the risk bound are based on relevant details of chromosomal abnormalities and MCC available in [4.1].

In essence the work reported in [4.1] gives on a single risk factor value, whereas the present study gives a more realistic statistical span of risk-factor.

CHAPTER V

STATISTICAL MIXTURE MODEL OF NORMAL AND ABNORMAL CHROMOSOMAL ADMIXTURE: APPLICATION TO CANCER GROWTH MODELS

5.1 Introduction

In the contexts of biostatistics, randomly-mixed state of two or more entities is generally indicated as a simple, prorated ratio of each existing population as described in the previous chapter. However, such arithmetic proportion could only be a limiting case of a "truly statistical mixture" as observed by Lichtenecker and Rother in [5.1] and by Neelakanta *et al*. in [5.2]. Though not specifically applied in biological contexts, relevant statistical mixture formulations of [5.1 and 5.2] can be considered as judicious candidates for example, to quantify the state of chromosomal mixture constituents in a cytogenic complex. Hence, considered in this chapter is a way to model the quantitative extents of normal and aberrated chromosomes that exist as a statistical mixture in a cytogenetic complex (in contrast with simple proportion model considered earlier in assessing such mixtures). Hence discussed in this chapter, are the following research efforts:

- To review the unique contexts of statistically-mixed constituents (at cytogenetic level) and apply statistical mixture theory to quantify their proportional existence
- To gather from literature relevant details on typical cytogenetic mixture constituents, their types, their population and contexts pertinent to normal and pathogenic conditions of certain genetic disorders

- Modeling the dynamics of growth (or decay) of a specific constituent (such as aberrated chromosomes) coexisting with other constituents like normal chromosomes as a part of the statistical mixture
- Determining the extent of chromosomal aberrations *via* statistical mixture theory *vis-a-vis* relevant etiology and genetic pathological states induced for example, spontaneous abortions
- Studying the competitive growth and decay dynamics of normal and abnormal chromosomes coexisting in a cytogenetic complex leading to the proliferation of observed pathology like cancer symptoms.

Necessary introductions on the above considerations and the associated details are outlined below:

5.1.1 Proportional-content Theoretics Applied to Cytogenetic Constituents: A Revisit

As discussed in the previous chapter, the state of chromosomal constituents has been studied classically, in terms of fractional population of normal chromosomes *versus* other cell contaminations. For example, as detailed in the previous chapter, studied in [4.1] is a mathematical model where the maternal cell contamination (MCC) is estimated in terms of various embryonic chromosomal attributes. In general, the fraction of cell contamination is termed as cytogenetic factor; and, the net effect of such contamination is addressed in [4.1] *via* traditional statistics of expected average of the contents involved; and, pertinent study is performed using the classical approach *via* statistical analysis and χ^2–method. Hence, the significance of such analysis is indicated towards prenatal selections in spontaneous abortions caused by abnormal chromosomal contents. Hence, the simple proportion mixture-theory applied to normal and abnormal chromosomal contents at

embryonic level is used to deduce the gross abortion conditions in subjects at clinical levels. Apart from such proportional mixture statistics considerations studied in last chapter, a statistical-mixture theoretic framework can also be developed to address the cytogenetic contaminated states. Relevant details and implications are as follows.

5.2 A Review on Statistical Mixture Theory: Applications to Cytogenetic Contexts

At the cellular level, the inherent aspects of cytogenetic complex correspond to a set of encoded messages associated with the structural and functional attributes of chromosomes. Each non-aberrant chromosome occurs at a specified probability in the cytogenetic complex in conformance with the associated cytogenetic information. Pertinent to such informatic statistics of chromosomes, one can associate an efficiency factor, related to the constituents of the cytogenetic complex having aberrant chromosomes; and, the extent of abnormality can then be optimally deduced in terms of this information efficiency factor.

5.2.1 Information efficiency (η) pertinent to Cytogenetic Informatics

In order to define and evaluate the cytogentic information efficiency (η), an universe (Ω) of a complex system made of the mixture entities, namely normal and abnormal chromosomes can considered with the total chromosomal population being N.

Suppose a constant c_i is assigned as a cost-factor to each chromosome (indexed as i = 1, 2…., N) whose occurrence probability is $P_{\ell i}$. Then, the average cost per chromosome (cost-function) can be written as follows:

$$C_{av} = \sum_{i = 1, N} P_{\ell i} c_i \tag{5.1}$$

The optimization of this cost-function (in constituting the overall cytogenetic layout) would refer to a value of $C_{av} \equiv \bar{C}$ over the entire set of chromosomes and subject to certain constraints on the associated entropy (Shannon information) profile. The lowest bound (*infimum*) of $\bar{C}$ is then given by [5.3],

$$\bar{C}|_{inf} = H(x)/\ln(N) \tag{5.2}$$

where N as indicated earlier, depicts the total chromosomal counts and H(x) is the entropy of the ensemble of aberrated chromosomes being present. The constrained optimization of interest is to optimize the cost-function as above in order to determine the net effect of coexisting aberrant chromosomes along with normal chromosomes. That is, inasmuch as both normal and aberrant chromosomes prevail in the cellular complex as mixture constituents, the entropy of the ensemble of this mixture should be viewed in terms of the redundancy arising from the prevalence of the aberrant chromosomes (contributing negentropy) in the system and implicating the said cost-function.

Hence, it is possible to define an information efficiency (η) factor of the cytogenetic complex using the classical concepts of information theory. It is the ratio of the average information (per chromosome) of the ensemble to the maximum possible (average) information (per chromosome) [5.3]. That is,

$$\eta = H(x)/\bar{C}\ln(N) \tag{5.3}$$

And, concurrently, $(1 - \eta)$ can be regarded as a redundancy factor (R). It refers to the reduction in information content of an ensemble from the maximum possible and it is specified as follows:

$$R = (1 - H/H_{max}) \tag{5.4}$$

where, the entropy functional denotes any H statistical divergence (or distance) metric such as, Jensen-Shannon (JS), Kullback-Leibler measure etc.[5.4]

5.2.2 Quantifying the Mixture Attributes of Cytogenetic Complexity via Redundancy factor

The aforesaid redundancy factor (R) can be attributed to a cytogenetic complex *via* statistical mixture theory as follows: Following the concept of statistical mixture theory due to Lichtenecker and Rother [5.5], the underlying heuristics specifies a weighted probability r that describes the effective statistical attribute of the mixture proportioned by the attributes θ and $(1 - \theta)$. In terms of a binary mixture of two constituents 1 and 2 with populations n_1 and n_2, $\theta = n_1/(n_1 + n_2)$ and $(1 - \theta) = n_2/(n_1 + n_2)$. Relevant weighted probability is given by:

$$r(\theta) = P_{1\ell}{}^{\theta} P_{2\ell}{}^{1-\theta} \tag{5.5}$$

which is valid, within the statistical upper and lower bounds, namely, $(r_{min} \leq r \leq r_{max})$. Explicitly, r_{min} and r_{max} are given by:

$$r_{max} = \theta P_{1\ell} + (1-\theta) P_{2\ell} \tag{5.6a}$$

and

$$r_{min} = \left[\frac{\theta}{P_{1\ell}} + \frac{(1-\theta)}{P_{2\ell}} \right]^{-1} \tag{5.6b}$$

The statistical bounds as above conform to the extreme arithmetic and geometrical-mean statistics of the constituents. With reference to the set $\{r: r_{min} \text{ and } r_{max}\}$, the corresponding Shannon measure of entropy (negentropy), I can be written as a function of r as follows:

$$I(r) = -H(r) = -r\ln(r) \tag{5.7a}$$

$$I(r_{min}) = -H(r_{min}) = -r_{min}\ln(r_{min}) \tag{5.7b}$$

$$I(r_{max}) = -H(r_{max}) = -r_{max}\ln(r_{max}) \tag{5.7c}$$

Suppose one of the constituent entities of the statistical mixture, say, the one with a population n_2 has a uniform distribution implying that the occurrences of its elements (in the statistical mixture space) are equally-likely. That is, ($P_{21} = P_{22} = P_{23} = \ldots = P_{2n2} = 1/n_2$), so that, [$P_{21} + P_{22} + P_{23} + \ldots + P_{2n2} = 1$]. In contrast, the other constitutive entity (with a population n_1) is presumed to be of elements each bearing a distinct probability of occurrence. That is, ($P_{11} \neq P_{12} \neq P_{13} \neq \ldots \neq P_{1n1}$) and [$P_{11} + P_{12} + P_{13} + \ldots + P_{1n1} = 1$]. In the context of cytogenetic complex, the population of aberrated chromosomes can be regarded as that with no information and as such, it belongs to the subset of cardinality, n_2 with considerations of equally-likely occurrences. This uniformly-distributed entities assumed as above is consistent with the so-called Laplacian concept on probability of equally-likely occurrences. For example, aberrated chromosomes coexisting with normal chromosomes in a mixed state within a cellular system denote such population n_2, inasmuch as the roles of such aberrated chromosomes are non-informative in the regular functions of the cellular complex. The presence of such aberrated chromosomes denotes a state of maximum entropy resulting from its uniformly-distributed contents. On the contrary, pertinent to each normal chromosome, it prevails with a distinct (unequal) occurrence probability as decided by its designated structural and functional characteristics. Hence, normal chromosomes can be regarded as informative

(negentropic) entities of the cytogenetic complex (However, the informative chromosomes could be redundantly present).

From information theory point of view, it is known that equally-likely occurrences of entities or events of a set of random variants mean a degree of certainty whereas, random (unequal) chances of occurrences imply an associated uncertainty of the set. The certainty consideration will bring down the negative entropy (or information content) of the set while, any uncertainty involved will augment the negative entropy.

Thus, the existence of equally-likely probabilities associated with the elements (such as aberrated chromosomes) of a mixture, it amounts to specifying an efficiency to the associated information content of the whole set; and, the related considerations lead to a redundant information-theoretic attribute [5.3] to the set in Shannon's sense. Such a measure of redundancy (R) can be specified in the context of a statistical mixture as indicated earlier by (equation 5.4):

$$R = 1 - \frac{H(r)}{[H(r)]_M} \tag{5.8a}$$

where $[H(r)]_M$ denotes the maximum value of H(r) over the fraction $0 \leq \theta \leq 1$ (or $1 \geq (1 - \theta) \geq 0$) of the binary mixture constituents. Further, referring to the upper and lower bounds on r specified by equation (5.6), the corresponding range of R can be deduced as follows:

$$R_{min} = 1 - \frac{H(r_{max})}{[H(r_{max})]_M} \tag{5.8b}$$

and

$$R_{max} = 1 - \frac{H(r_{min})}{[H(r_{min})]_M} \tag{5.8c}$$

Thus, the complexity metric of a statistical measure evaluated using the redundancy measure (R) as above can be adopted to analyze the entropy features of the binary constituents of such mixtures. That is, considering a heterogeneous mixture of normal and aberrant chromosomes, their relativeness in causing genetic disorders can be estimated using the parameter, R.

5.3 Statistical Mixture of Cytogenetic Contents: Implications on Observed Pathogenic States

In view of the general considerations in describing the statistically-mixed state of complexity of normal and abnormal chromosomes outlined above, the scope of the present study is to address the following specific tasks on certain clinically observed pathogenic conditions (such as cancer growth).

- Copy-number alterations *versus* cancer growth
- Ploidy and aneuploidy involvement in oncological contexts
- Clonal and non-clonal alterations implicating cancerous growths.

The following sections are devoted to describe the underlying considerations and hence, proposed are algorithms/computations relevant to cancer growth dynamics etc.

5.4 Genetic Disorder and Cancer-Growth Considerations: Modeling *via* Copy number Alterations

As indicated in Chapter 3, the copy-number and its alterations refer to changes in DNA of a genome that result in the cell having an abnormal number of copies of one or more sections of the DNA. One of the genetic aspects of cancer results from irreversible

structural mutations manifesting as changes in DNA copy number at distinct location in the genome [5.7]. Aberrations of this type affect the functions of a gene; and, in general, understanding and quantifying such aberrations is essential to comprehend the disease etiology. This could possibly help developing targeted therapies in gene-related pathogenic states.

Developed in [5.7] is a multi-component scoring model for copy-number alterations (CNA) that cause genetic defects. The presence of such undesirable entities (CNA) is modeled as "noise" and the associated tumor heterogeneity is described in terms of related "noisy impurities" designated as 'stromal admixture constituent'. Typically four scoring parameters are identified in [5.7] to quantify the copy-number alterations. These are as follows: Single-copy gain (A_o), amplification (A_1), hemi-zygomatic loss (D_o) and homozygous deletion (D_1). The explicit definitions of A_o, A_1, D_o and D_1 are as follows:

Single-copy gain (A_o): When a gene copies itself, and the repeats are located in small clusters (known as tandem repeats) or spread throughout the genome, it is defined as single-copy gain

Amplification (A_1): This refers to selective, repeated replication of a certain gene or genes without a proportional increase in other genes in the genome

Hemi-zygomatic loss (D_o): Hemi-zygous means having one copy of a gene instead of usual two copies. For example, male is hemi-zygous for most X chromosome genes and any loss occurred in above condition is known as hemi-zygomatic loss

Homozygous deletion (D_1): Homozygous means having two identical alleles for a given trait; and, deletions are fragments of chromosomes that are missing.

5.5 Ploidy and Aneuploidy Involvement in Tumor Growth

In the contexts of observed carcinoma, the non-aberrant cells may also coexist with the aberrant cells. In such situations as mentioned in earlier (Chapter 3), the *aneuploidy* refers to the presence of abnormal chromosomes in a cell and it is an indication of chromosomal abnormality. Aneuploidy could imply either missing or the presence of extra chromosome states within the cytogenetic complex causing genetic disorders with relevance to some forms of cancer.

The fraction of ploidy and aneuploidy states for example, may estimate the extent of tumor growth as indicated in [5.8] with reference to breast carcinoma. The infiltration aneuploidy into the non-aberrant cells would result in tumor conditions across different extents of ploidy, from haploid to polyploidy levels. Given a ploidy, the tumor is observed when a corresponding aberrant cell fraction exists. Typically for low ploidy, high fractions are indicated to confirm tumor conditions. With higher ploidy, however tumor can be seen even at lower aberrant cell fraction. An exercise called *allele-specific copy number analysis of tumors* (ASCAT) first determines the ploidy of tumor cells and specifies the fraction of aberrant cells.

5.6 A Mixture-state Model of Cytogenetic Complex in Terms of Clonal and Non- clonal Alterations

In cytogenetic contexts, the abnormal number of chromosomes is of interest in developing models that represent the dynamics of cancer evolution resulting from patterns of chromosomal aberrations [5.9]. Relevant chromosome abnormality is specified in [5.9] in terms of clonal chromosomal aberrations (CCA) and non-clonal

chromosomal aberrations (NCCA). (The definitions of CCA and NCCA are given in Chapter 3). The NCCAs include numerical changes (aneuploidy and structural aberrations). When the aberration observed due to translocation is less than 20%, then it is considered as NCCA; otherwise, it is CCA.

The genesis and growth of cancer can be specified by a kinetic model in terms of proportions of NCCA and CCA. Presented in [5.9] is a kinetic model of cancer evolution and progression elucidated in terms of four entities and related pathway considerations. The interacting dynamics of normal cell (A), NCCA (N) and CCA (C) plus the associated pathways of interaction are illustrated in Figure 5.1.

In the model of Figure 5.1, there are interacting coefficients denoted as k_1, k_2…, k_6. Further, the set $\{k_1, k_2\}$ imply balancing trend imposed towards depletion of A, if balancing trend is positive; otherwise A is subjected to repair and application; likewise, k_3 and k_4 represent the balance between the depletion of CCA (C) and NCCA (N) respectively if positive, and, if negative, it implies the repair and replication of CCA and NCCA. The set k_5 and k_6 denote the balance between the demise of NCCA and CCA respectively, if positive; and, when negative, it implies the repair and replication of NCCA and CCA. Further, the four quantities A, N, C and D (denoting the concentration of nonfunctional chromosomes) can be rendered as dynamic variables changing with respect to time (τ), further at any instant (τ), $A(\tau) + N(\tau) + C(\tau) = 1$.

Hence, it is assumed in [5.9] that A, N and C are growth functions (for example, exponential functions of time with a time-constant) and relevantly, the resulting cancer growth function is indicated in terms of the functional attributes of A, T and C with respect to time in the context of tumor proliferation as described below.

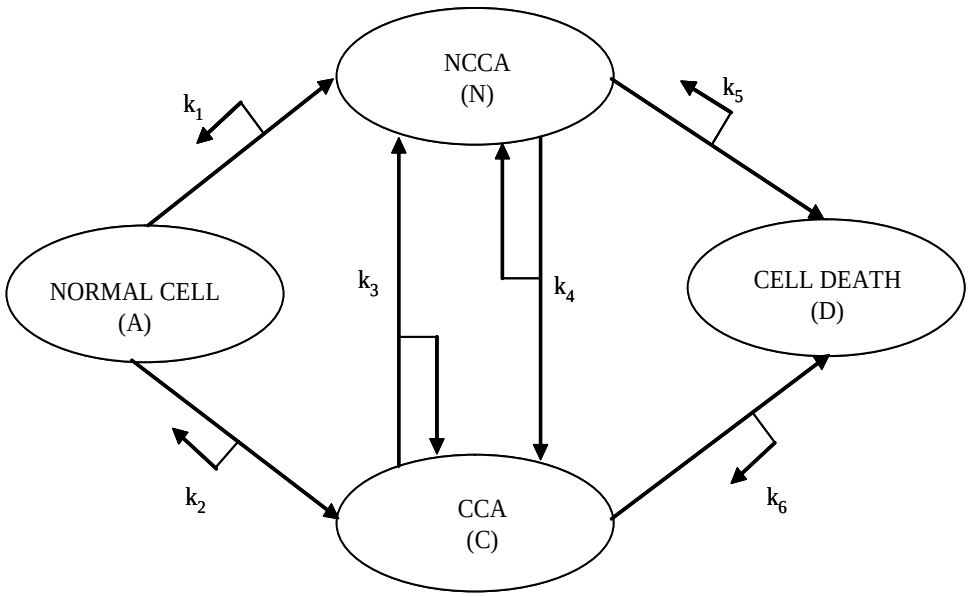

Figure 5.1 Proposed pathway for cancer evolution and progression when k_n is equal to the pseudo reaction rate constant [5.9]

5.7 Dynamics of CCA and NCCA Profiles

As described in the previous section, a cell depicts a cytogenetic complex system containing CCA and NCCA entities being present in a state of statistical mixture. Relevantly, a complex system model can be ascribed to specify the associated spatiotemporal states as outlined in the following subsection. Further, in the complex system framework of a cell, the dynamics of the associated activity as regard to the growth or decay of the constituent populations can be modeled in terms of a stochastical differential equation as presented later.

5.7.1 Quantifying Stochastical Mixture Attributes of the Contents in the Cytogenetic Complexity [5.3]

Consider the universe of a complex system specified by a domain $\mathbf{\Omega} \in \{\mathbf{X}(x)\}$ as illustrated in Figure 5.2. Suppose two constituent (interactive) subsystems $\{x\ (\mu);\ i = 1, 2,$

..., ℓ, ..., n_1} and {x (ν); j = 1, 2, ..., ℓ,..., n_2} are respectively characterized by two sets of attributes μ and ν; that is, **Ω** {x (μ, ν)} where x ∈ **X** and $(n_1 + n_2) = N$ depicts the cardinality of the total universe **Ω** of the compositional domains.

Further, the occurrence probabilities of the sets {x (μ); i = 1, 2, …, ℓ, …, n_1} and {x (ν); j = 1, 2, …, ℓ,…, n_2} are $\{P_{1\ell}\}_{\ell = i}$ and $\{P_{2\ell}\}_{\ell = j}$ with the subscripts 1 and 2 depicting the attribute sets {μ}and {ν} respectively.

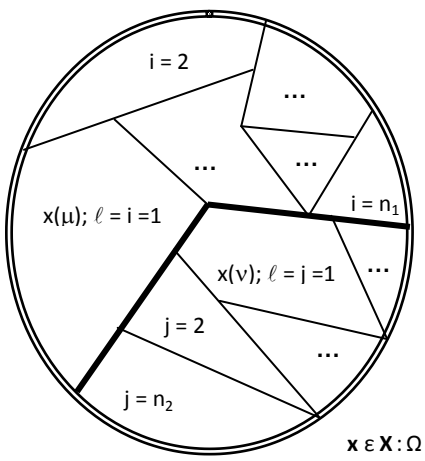

Figure 5.2 A complex system **Ω**: **X** depicting a mosaic of statistical mixture constituted by a pair of binary subsystems (compositional domains) [5.3]

Suppose the randomness associated with the subsets of Figure x is expressed in terms of the aforesaid occurrence probabilities $P_{1\ell}$ (μ; i : n_1) and $P_{2\ell}$ (ν; j : n_2), (corresponding to the attribute sets {μ} and {ν}, respectively). Now, the maximum entropy concept [5.10−5.12] applied to each group in the domain **X** leads to the following entropy functionals:

$$H(s_\mu) = \ln(n_1 + 1) \approx \ln(n_1) \text{ with } n_1 >> 1 \qquad (5.9a)$$

$$H(s_\nu) = \ln(n_2 + 1) \approx \ln(n_2) \text{ with } n_2 >> 1 \tag{5.9b}$$

where s_μ and s_ν refer to some metrics of gross complexity corresponding to the extensiveness of the populations of the sets$\{x(\mu); i = 1, 2, \ldots, n_1\}$ and $\{x(\nu); j = 1, 2, \ldots, n_2\}$respectively.

With reference to a complex system viewed in an entropy-based framework, the global complexity (depicting s_μ and s_ν) has been described in [5.10−5.12] in terms of a complexity metric (S). It is defined using the associated disordered sets of constituents and it corresponds to a solution equal to $\exp(-\beta)$ where β is a Lagrangian that maximizes the entropy functional of the complex system. Further, considering a large set of disordered entities (constituting a complex system), s defines a dichotomy of two regimes [5.10]: (i) $0 \leq S < 1$and (ii) $1 < S < \infty$. When S is very small ($S \to 0$), the system is regarded as "simple"; and, as $S \to \infty$, the system becomes totally complex. (The value of $S = 1$ is a transition that bifurcates the system of being simple or complex when viewed in terms of the entropy involved).

Equations (5.9a) and (5.9b) are consistent with the so-called Jaynes' principle of maximum entropy or maximum uncertainty and a class of distribution corresponding to the maximum entropy formalism has been identified in [5.13] to exist. Further, equation (5.9a and 5.9b) concurrently leads to the following Shannon information formulations [5.14]:

$$I_1 = \{x(\mu); i = 1,2,\ldots,n_1\} = -\sum_{x \in X; \ell = i} P_{1\ell} \ln(P_{1\ell}) \tag{5.9c}$$

and

$$I_2 = \{x(\nu); j = 1,2,\ldots,n_2\} = -\sum_{x \in X; \ell = j} P_{2\ell} \ln(P_{2\ell}) \tag{5.9d}$$

Equations (5.9c) and (5.9d) can be regarded as implicit representations of gross complexity pertinent to the sets {x (μ); i = 1, 2,…, n_1} and {x (ν): j = 1, 2, …, n_2} respectively, *in lieu* of the relations specified by equations (5.9a) and (5.9b). While equation (5.9a) depicts the maximum entropy measuring the gross complexity (s) of the set {x (μ); i} or {x (ν); j}, an alternative metric can also be specified to measure the relative complexity between these sets. It refers to a pair of cross-entropy functionals, which can be written in the following forms [5.12]:

$$H(s_\mu \,||\, s_\nu) = D(s_\mu \,||\, s_\nu) = \sum_{\ell = i} P_{1\ell} \ln(P_{1\ell}/P_{2\ell}) \tag{5.10}$$

$$H(s_\nu \,||\, s_\mu) = D(s_{\nu|} \,|\, s_\mu) = \sum_{\ell = j} P_{2\ell} \ln(P_{2\ell}/P_{1\ell}) \tag{5.11}$$

The cross-entropy functionals of equations (5.10) and (5.11) denote synonymously the "statistical divergence" D ($s_\mu || s_\nu$) between the random attributes of {x (i): μ} *versus* {x (j): ν}, or *vice versa.* This cross-entropy measure also refers to relative or mutual information content in Shannon's sense. Further, the measure specified *via* equations (5.10) and (5.11) follows Kullback's minimum (directed) divergence or minimum cross-entropy principle [5.15].

In addition, the cross-entropy concept of depicting the relative complexity as above, implicitly implies an expected logarithm of the likelihood ratio (L), namely,

$$L = \frac{[(P_{1\ell})_{\ell = i}]_I}{[(P_{2\ell})_{\ell = j}]_{II}} \tag{5.12}$$

where $[P_{1\ell}]_I$ and $[P_{2\ell}]_{II}$ are respective probabilities of observations of the attributes $\{\mu\}$ and $\{\nu\}$ in the complex system when a certain hypothesis (h_I, h_{II}) is true. Corresponding to L, a *log-likelihood ratio function* (LLR) given by ln(L), can be defined and regarded as a "discrimination measure" that provides a choice, whether to choose $\{\mu\}$ in preference to $\{\nu\}$ or *vice versa.* (The LLR is well-known [5.12] as a useful metric in decision-making efforts and can be considered identically to depict a measure of contrast between the constituents involved).

Designated as *Jensen-Shannon* (JS) *measure* [5.16], it is a variation of the Kullback-Leibler divergence formulation and it is explicitly given by the following expression:

$$JS_{\pi}(P_{1\ell}, P_{2\ell}) = H(\pi_1 P_{1\ell} + \pi_2 P_{2\ell}) - \pi_1 H(P_{1\ell}) - \pi_2 H(P_{2\ell}) \tag{5.13a}$$

where $(\pi_1, \pi_2) \geq 0$ and $(\pi_1 + \pi_2) = 1$; and,

$$H(P_{1\ell}) = -\sum_{\ell=i} P_{1\ell} \ln(P_{1\ell}) \tag{5.13b}$$

$$H(P_{2\ell}) = -\sum_{\ell=j} P_{2\ell} \ln(P_{2\ell}) \tag{5.13c}$$

$$H(\pi_1 P_{1\ell} + \pi_2 P_{2\ell}) = -\sum_{u=i} \sum_{v=j} (\pi_1 P_{1u} + \pi_2 P_{2u}) \ln(\pi_1 P_{1v} + \pi_2 P_{2v}) \tag{5.13d}$$

The weights π_1 and π_2 in the above relations for example, can be taken respectively as $\theta = n_1 / (n_1 + n_2)$ and $(1 - \theta) = n_2 / (n_1 + n_2)$ in the context of a mixture complex.

Exclusive to a binary statistical mixture, a measure of global complexity can be specified in terms of the maximum entropy associated with the disordered constituent

entities (1 and 2) (assuming each having a large population namely, (n_1 and n_2) $\rightarrow \infty$ and exist as a mixture of specified proportion).

5.8 Temporal Changes in the Cytogenetic Constituents: Neoplasmic evolution

The cellular contents at the cytogenetic level expressed in terms of quantitative chromosomal aberrations are important in deducing certain genetic-based diseases such as, carcinogenesis. Hence, mathematical models have been developed to indicate the extent of cancer growth that is, the neoplasmic evolution with respect to time (t). For example, described in [5.17], is the progression of human cancer characterized by the accumulation of genetic instability due to chromosomal aberrations; and, it is indicated that the chromosomal abnormalities plus the aneuploidy could be as high as 90% in human malignant tumors. Relevantly, it is specified that, the underlying idiographic features of tumors can be analyzed *via* cytogenetic and/or molecular features pertinent to chromosomal entities.

Modeling of tumor development in general, can help understanding carcinogenesis and the related dynamics of pathogenic conditions. Hence, basic linear models were classically developed for example, as in [5.18] for colorectal tumors; however, the inter-relationship between chromosomal abnormalities coexisting in the cytogenetic complex may force the underlying growth model to be nonlinear. In addition, a biologically comprehensive neoplastic development should be viewed in a stochastical framework (rather in terms of deterministic variables) consistent with the complexity of the cellular system and its contents having spatiotemporal randomness

An effort towards addressing the balanced/unbalanced states of cytogenetic contents, which take into relevant interaction considerations alone, has been exercised by a

stepped progression of the *number of imbalances of the tumor* (NIBT) [5.19]. It denotes a discrete version of nonlinear evolution of cancer growth.

More rigorously, a continuous nonlinear mathematical model in the context of interaction between tumor cells and oncolytic viruses has been developed in [5.20]; and, with the help of differential-equation calculus, relevant stability of the system has been established. In essence, the model considers infected and uninfected versions of tumor cells growing in logistic fashion. The stochastical aspect of tumor invasion in the surrounding space degrading the extra-cellular matrix is studied *via* two classical efforts due to Anderson and Chaplin [5.21] and Othmer and Stevens [5.22].

The Anderson-Chaplin model [5.21] is framed on the basis of various biological and biomedical considerations; and, hence the formation of endothelial cell (EC) surface developing in response to certain chemicals (called TAF) is described *via* EC migration of spatio-temporal dynamics using ordinary differential equations (ODE). Additional random features to the said dynamics are included in the model due to Othmer and Stevens [5.22].

More comprehensively, developed in [5.23] is an asymptotic profile of the solution to parabolic ordinary differential equations pertinent to tumor angiogenesis dynamics. It resolves thereof certain solvability issues of ODEs seen in [5.21] and [5.22].

Yet another model on early tumor growth and invasion has been developed *via* cellular automaton considerations by Patel et al. [5.24]. Essentially, a hybrid cellular automaton model of early tumor growth that describes the activity of individual cells and continuous evolution and their microenvironment, forms the main theme in [5.24].

5.8.1 Chromosomal Aberrations: A Stochastical Profile

An important query of interest in modern context is to know whether the chromosome aberrations are random events or have arised from an internal (endogenous) deterministic mechanism. Discussed in [5.25] is the stochastical nature of chromosome aberrations in solid-tumors; and, several related Shannon information functions are evaluated thereof to describe the disorderliness present inside a tumor. Hence, suggested in [5.25] is that, in the context of quantifying the spread of aberrations, the generating process is neither deterministic nor totally random; but, it produces variations that can be specified between two extrema. The study in [5.25] is fortified with relevant data on 79 different kinds of solid-tumors having 30 or more karyotypes retrieved from [5.26].

The mixture-state of contents expressed in terms of the stains at DNA level in the chromosomal structures is featured and expressed in terms of autosomal genetic markers in [5.27]. Specifically the combinations of alleles at different locales (loci) on the chromosome that are transmitted together are considered. This one fold, single and simple allele (known as *haplotype)* forms the general framework to test the hypothesis on mixed stain analysis advocated in [5.27].

The implication of neo-plastic transformational dynamics concerning the induction of chromosomal aberration through direct and bi-stander mechanism has been addressed *via* a state-vector model.

5.8.2 Stochastical Dynamics and Bernoulli-Riccati Equation on Tumor Growth Model

Notwithstanding various growth models of tumors due to chromosomal aberrations, yet considered here is a method to predict a compactable growth function that describes the dynamics of evolution of tumor *via* the presence of disordered entities constituted by a

stochastical mixture of normal and abnormal chromosomes present in the receptacle matrix of the body, where the tumor is observed. In all, the spatio-temporal framework of the disordered set of entities mentioned above is assumed as a complex system and hence relevant stochastical time-evolution of tumor growth is indicated in closed functional form. This growth function is verified against the results obtained by other models.

In the context of a complex system description pertinent to the cytogenetic framework, the associated growth or decay of the constituents (being present as a stochastical mixture) can be elucidated *via* stochastical differential calculus as described below.

Pertinent to the nonlinear growth function under consideration, by resorting to a natural extension of the first-order equation in calculus, namely $y'(t) = p_o(t) + q_o(t)$, a deterministic nonlinear, first order equation of the following type can be specified:

$$y'(t) = p_o(t) + q_o(t)\, y + r_o(t)\, y^n \tag{5.14}$$

where $p_o(t)$, $q_o(t)$ and $r_o(t)$ are continuous functions of t and $n \neq 0$. The above equation is well-known as the *Bernoulli equation* which can be transformed into an integrable form using an appropriately chosen (new) dependent variable.

In equation (5.14) the functional coefficients p_o, q_o and r_o denote definite attributes of the nonlinear activity in the deterministic framework as governed by the differential equation. The coefficient p_o characterizes the external stimulus that enables the nonlinear activity to commence and remain sustained. Depending on the output value y, the prevailing activity is weighted by the extent of that output value leading to the nonlinearity perceived. This is accommodated in equation (5.14) by the coefficient r_o, which decides the

output-dictated influence on the nonlinear activity. (Should r_o be equal to zero, the activity degenerates to a simple linear input-output relation). The exponent n stipulates the degree of the output entity in formulating the extent of nonlinearity involved. If n = 2, equation (5.14) is known popularly as the *generalized Riccati equation*. The coefficient q_o in equation (5.14) contributes to the linear input-output relation and could be set equal to zero, if the nonlinear activity predominates. It should be noted that in general, p_o, q_o and r_o could be constant coefficients as well.

Assuming the extent of nonlinearity limited by n = 2(first order nonlinearity), the Bernoulli equation given by equation (5.14), can be reduced to a simpler Riccati equation namely,

$$y'(t) + r_o(t)y^2(t) + p_o(t) = 0 \tag{5.15}$$

Equation (5.15) assumes that the underlying activity is more likely (or predominantly) nonlinear and the output y(t) that governs such a nonlinear behavior. It implies a simple case with a second degree influence (with n = 2). Further, it is indicated in [5.11] that the solution to equation (5.12) can be written as follows:

$$y(t) = L_Q(t) \tag{5.16}$$

where $L_Q(t)$ is popularly known as the *Bernoulli or modified Langevin function*. (It is also, sometimes referred to as *Brillouin function or Langevin-Bernoulli function*). The function $L_Q(t)$ is explicitly given by: $(1+1/Q) \coth [(1+1/Q)\, t] - (1/Q) \coth [1/Q)\, t]$.

When indicated for a nonlinear stochastical process, the Q-value in equation 5.16 decides the extent of disorderliness associated with the system. For example, in statistical mechanics [5.11], Q denotes an “order parameter” for the underlying stochastical process;

and, when Q → ½, it refers to a total isotropic disorder. When Q → ∞, the system is presumed to settle at a totally ordered-state. Hence, in the event that the underlying growth/decay process indicated in a stochastical framework, relevant dynamics can be stipulated within a pair of upper and lower bounds corresponding to Q → ½ and Q → ∞, respectively.

The application of the above model to the stochastical aspects of cancer growth [5.9] is presented in the following section (Result).

5.9 Cancer Growth Models

Presented in [5.9], a mathematical model relating chromosomal aberrations to cancer progression. As discussed earlier, the details in [5.9] are mathematical formulations on the dynamics of cancer in terms of the associated cellular immortalization process so as to get an insight into cancer initiation and progression useful for new therapies.

Table 5.1 Stage specific mutation pseudo reaction rate constants [5.9]

Stage	k_1	k_2	k_3	k_4	k_5	K_6
0.5	0.3570	1.7184	0.5829	2.0579	−2.1074	3.6334
1.5	0.2329	0.4361	0.6583	0.0043	−0.0733	−0.3289
2.5	−2.2540	0.0311	0.5793	−0.3865	0.2343	−1.2784
3.5	Undefined	Undefined	2.5799	0.2894	0.3028	−2.4589
4.5	Undefined	Undefined	−1.4807	1.4906	−1.1675	0.9254
5.5	Undefined	Undefined	−0.4786	0.4282	−3.1675	0.7142
6.5	Undefined	Undefined	0.0809	0.4338	−0.6497	0.1562

Further, the data adopted in [5.9] refers to experimental studies at the Center for Molecular Medicine and Genetics and Wayne State of University, School of Medicine. It represents stage visual snap-shots of the biological process for a specific cancer progression. It refers to wet-lab experiments on specimen harvested from p^{53} mutation leading to determining chromosome numbers and translocations classified into NCCA or CCA category.

Table 5.2 Calculation of C*, N* and CH*

Stage τ	C	N	CH	Sum (C+N+CH)	N*= N/Sum	C*= C/Sum	CH*= CH/Sum
0.5	0.2749	0.2741	1	1.549	0.17695	0.17746	0.64557
1.5	0.3039	0.5667	1	1.8706	0.30295	0.16246	0.53458
2.5	0.2129	0.8159	1	2.0288	0.40215	0.10493	0.49290
3.5	−0.0532	0.9816	1	1.9284	0.50902	−0.02758	0.51856
4.5	0.8489	0.5343	1	2.3832	0.22419	0.35620	0.41960
5.5	0.8477	0.1503	1	1.998	0.07522	0.42427	0.50050
6.5	0.7479	0.2403	1	1.9882	0.12086	0.37616	0.50296

Using the experimental data as above, empirical curves on the concentration ratio of normal chromosomes (A), concentration ratio of NCCA (N) and concentration ratio of CCA (C), are indicated as functions of cancer growth stages (τ) with, $A(\tau) + N(\tau) + C(\tau) = 1$.

For empirical depiction *via* least-square curve fitting, the growth and decay functions are assumed as corresponding exponential functions with appropriate coefficience and reaction rate-constants for the growth or decay. The co-efficients are taken in normalized set equal to 1 and the rate constants are indicated by k_1, k_2...,k_6 relevant to the following differential equations of the dynamics involved.

$$dA(\tau) / d\tau = -k_{12}A \tag{5.17}$$

$$dC(\tau) / d\tau = k_2A(\tau) + k_4N(\tau) - k_{36}C(\tau) \tag{5.18}$$

$$dN(\tau) / d\tau = k_1A(\tau) + k_3C(\tau) - k_{45}N(\tau) \tag{5.19}$$

where, $k_{12} = k_1 + k_2$, $k_{36} = k_3 + k_6$ and $k_{45} = k_4 + k_5$.

The rate-constants used are indicated in Table 5.1

Table 5.3 Experimental and empirical data on N(τ) and C(τ) of [5.9] and computed data of the present model

	Experimental Results		Theoretical Results		Present Study	
τ	N_{obs}	C_{obs}	N_{theor}	C_{theor}	N: $[1-L_Q(\tau)]$	C: $L_Q(\tau)$
0.5	0.2741	0.2749	0.2038	0.2901	0.905864	1.079419
1.5	0.5667	0.3039	0.4156	0.3081	0.067601	1.786625
2.6	0.8159	0.2129	0.8934	0.3324	0.003587	1.840631
3.5	0.9816	−0.0532	0.9947	−1.3798	0.00018	1.843505
4.5	0.5343	0.8489	−230.90	1.5700	8.99×10^{-6}	1.843650
5.5	0.1503	0.8477	−6734	0.6283	4.47×10^{-7}	1.843657
6.5	0.2403	0.7479	0.8659	0.6383	2.23×10^{-8}	1.843658

The empirically fitted mathematical functions for A, C and N are as follows:

$$A(\tau) = A_O e^{-k12\tau} \tag{5.20}$$

where, $A_O = 0.99$ and $k_{12} = (k_1 + k_2)$

$$C(\tau) = k_4 / k_{346} (1 - e^{-k345\tau}) \tag{5.21}$$

$$N(\tau) = k_3 / k_{345} (1 - e^{-k345\tau}) + 0.01e-^{-k345\tau} \tag{5.22}$$

where, $k_{346} = (k_3 + k_4 + k_6)$ and $k_{345} = (k_3 + k_4 + k_5)$ and $A_O = 0$.

Present Model: Considered in this study is a non-empirical formulation for N and C variations with respect to τ. It is based on assuming a logistic growth function model of C or N with respect to the variable τ.

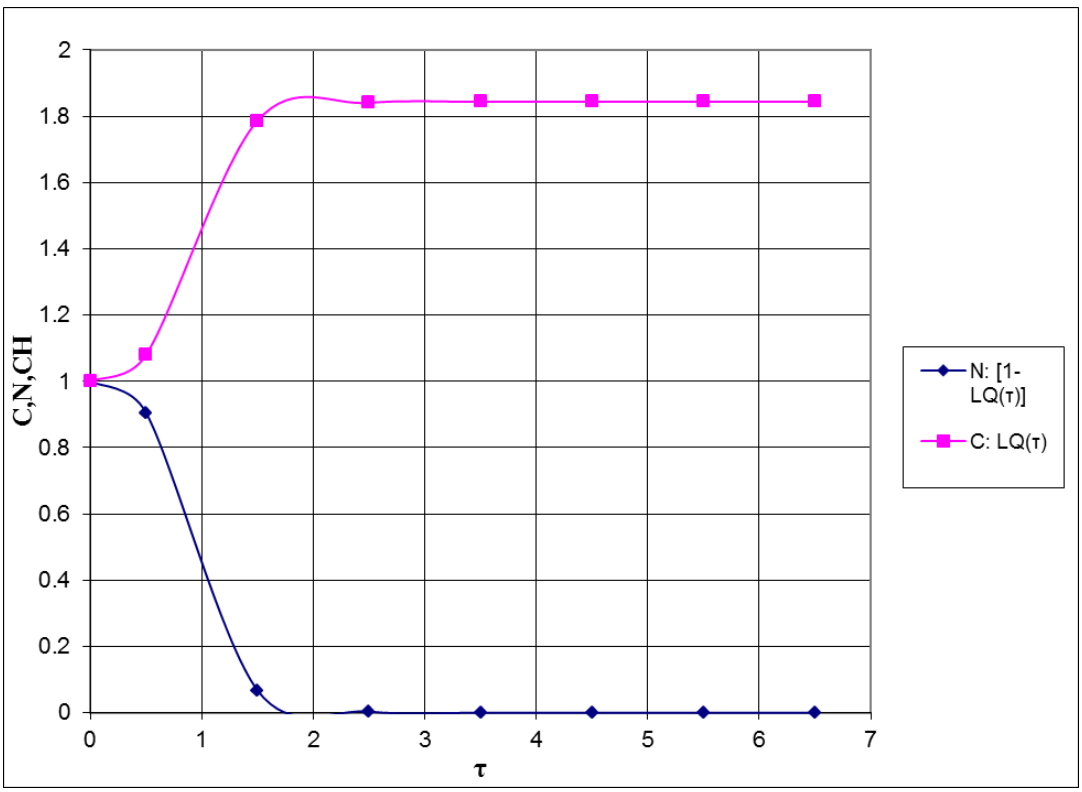

Figure 5.3 Plot of $L_Q(\tau)$ (series 1)and $1 - L_Q(\tau)$ (series 2) of Table 5.3

Inasmuch the growth of clonal and non-clonal constituents in the chromosomal complex is decided by the associated stochastical considerations of the populations of N and C and their interactions. In such complex system the use of Langevin-Bernoulli function has been indicated as feasible logistic function to denote the growth or decay profile of the constituent entities [5.28a & b].

That is, given a variable x, its growth as a function of time in normalized form, as $L_Q(t)$ and the corresponding decay function is $1 - L_Q(t)$. Here Q decides the order-disorder characteristics of the complex systems; and Q decides the initial rate (M_O) of growth or decay. The corresponding order-parameter of the system is given by P. S. Neelakanta and De Groff [5.29]. $\theta = 3/2\ ((M_O) - 1/2$, In terms of Q, $M_O = (1/3 + 1/3Q)$ and hence, $Q = 2\theta$. The order parameter θ can be assumed on equi-partition considerations equal to $(1/3)^{rd}$ of the constituent population namely, N*, C* and CH* (where the asterisk denotes the normalized values); and, CH denotes the total chromosomal content assumed as 100% denoting (clonal aberrations + non-clonal aberrations and normal chromosomes). The normalization indicated above refers to the sum of C, N and CH contents. For example, considering the growth stages of cancer indicated in [5.9], the following Table 5.2 is deduced.

Thus, $(\theta = N^* + C^* + CH^*)/3$ can be determined and $Q = 2\theta$ can be ascertain using this computed values of Q, $L_Q(\tau)$ and $1 - L_Q(\tau)$ can be determined. Therefore, the growth and decay values of C and N respectively are determined as function of τ as indicated in Table 5.3 along with the experimental data and empirical evaluations available in [5.9].

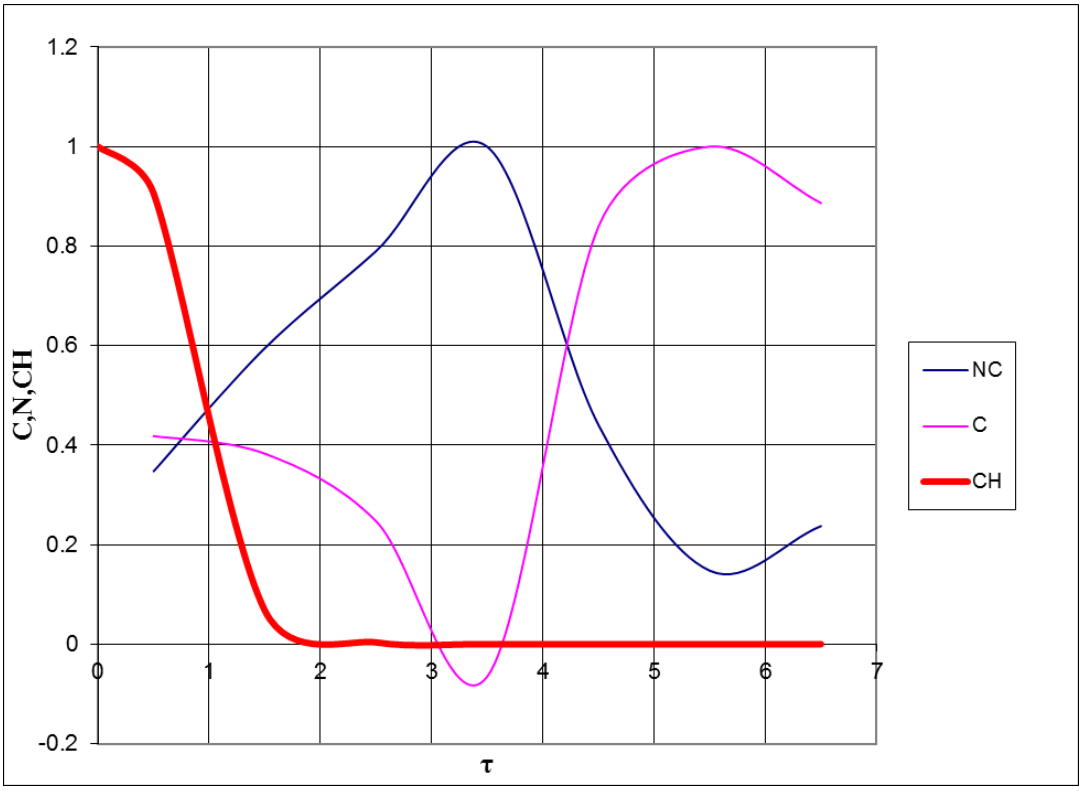

Figure 5.4 Experimental data in normalized form of C(τ) , N(τ) and CH (τ) and τ denotes the stage of cancer growth as available in [5.9]

The above results are graphically presented in Figures5.3, 5.4 and 5.5 respectively. For comparison, illustrated in Figure 5.4 are graphical representations of C(τ), N(τ) and CH(τ) relevant to (i) Experimental results [5.9], (ii) Empirical results [5.9] and (iii) Present study. Further, considering mathematical aspects of linking cancer and viruses, relevant cell responses are considered as modeling parameters of interest. Recent developments in genetic engineering aim at interaction between tumor cells and oncolytic viruses. Specifically as analyzed and presented in [5.20] is a mathematical model on such interaction dynamics and the associated stability consideration are obtained in relevant

analysis both infected and uninfected tumor cells being present is considered and the possibility of tumor load getting eliminated with time using virus therapy is suggested.

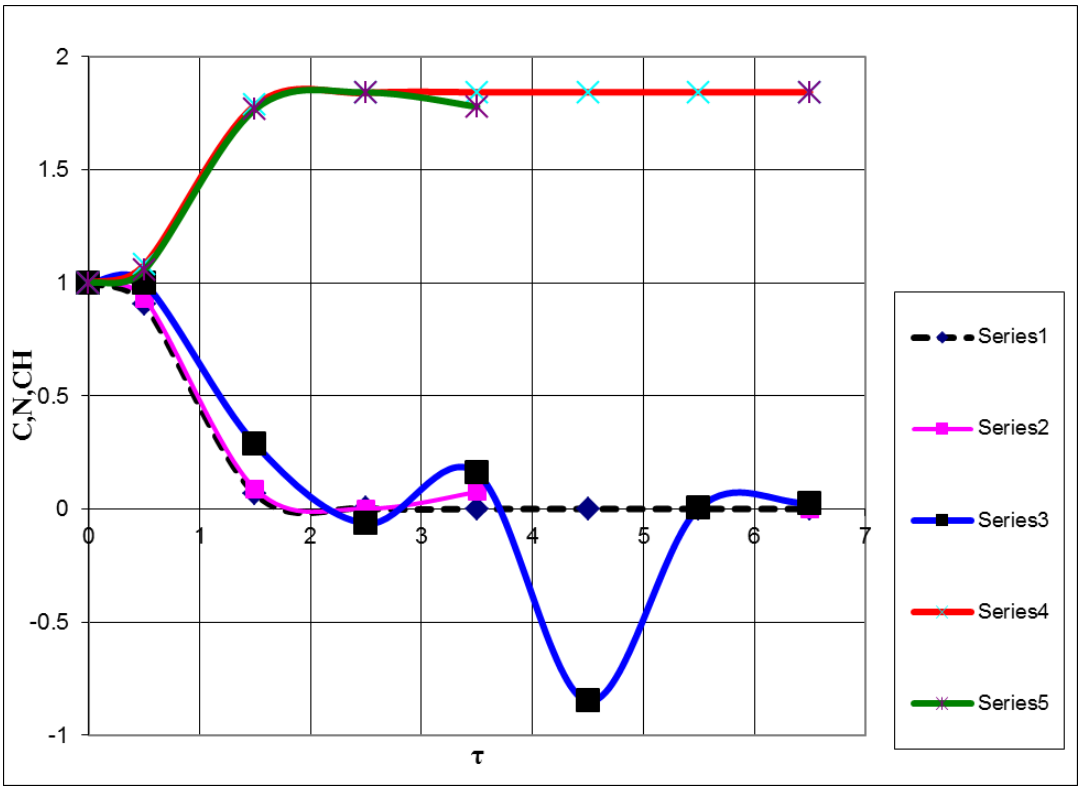

Figure 5.5 Plots of C(τ) ,N(τ) and CH(τ)

Series1: Decay (Present Method)
Series2: Decay [5.9]
Series3: Decay [5.9]
Series4: Growth (Present Method)
Series5: Growth [5.9]

In the model of [5.20] the tumor cells population representing as "X" and infected tumor cell population represented by "Y" are assumed to grow in logistic fashion. Further the following assumptions are made in the model proposed.

- The oncolytic viruses*(Oncolytic viruses*: These refer to virus that preferentially infect and lyse/break down the cancer cells) enter tumor cells and replicate subsequently
- The tumor cells get infected with invading oncolytic viruses
- The infected tumor cells then cause infection in other tumor cells.

The initial condition on X and Y are: $X(o) = X_o > 0$ and $Y(o) = Y_o > 0$.

Relevant bounded solutions on the cancer growth are obtained *via* differential equations formulated using the growth logistics. Relevant solutions for the differential equation are based on numerical procedure *via* Runge-Kutta method. Further, the computed result available in [5.20], conform to the following parameters:

- r_1 = Maximum growth rate of uninfected cell (= 40)
- r_2 = Maximum growth rate of infected cell (= 2)
- K = Holding or carrying capacity of the cellular media (= 100)
- a = A measure of immune response of the individual of the viruses that prevents it from destroying the cancer (= 0.05)
- b = Transmission rate of the viral dispersions (= 0.02)
- α = Rate of infected cell killing by the viruses (= 0.003).

Corresponding to the presumed data as above, the computed densities of tumor cells as functions of time are presented in [5.20] For the values of r_1 assumed, the initial growth rate seen in [5.20] is, m = 0.35 to 0.4. With value of b changed to 0.06, the initial slope m becomes approximately 0.7.

Given, the initial slope m and assuming a stochastical framework of growth/decay dynamics, suggested in this study, functions $L_Q(t)$ and the decay $(1- L_Q(t))$ to depict the growth and decay profiles respectively. Here, the Q-value denotes the order function (disorder factor), and it is related to m by the relation: $Q = 1/(3m-1)$. Hence when $m \approx 0.7$, $Q \rightarrow 1$ and when $m \approx 0.35$, $Q \rightarrow 20$. Thus, $(Q \rightarrow 1)$ data corresponding to $b = 0.06$ implies an upper-bound on growth dynamics; and $(Q \rightarrow 20)$ data for $b = 0.02$ specifies the lower-bound growth dynamics.

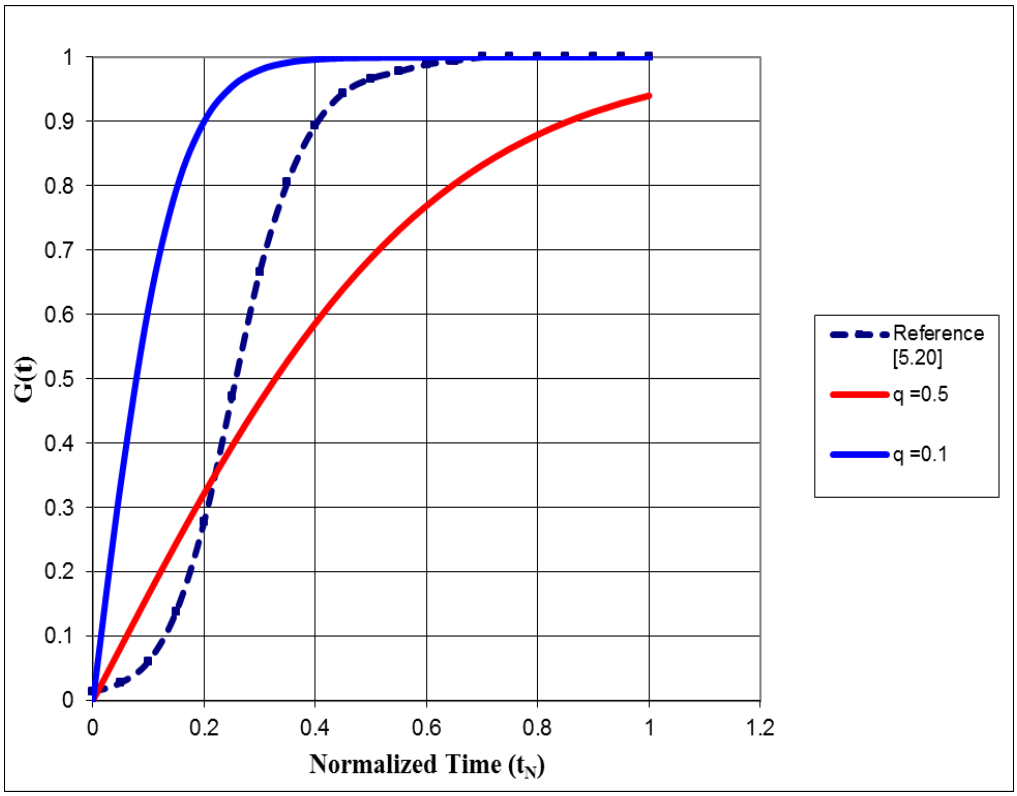

Figure 5.6 (a) Growth function curves

Shown in Figures 5.6 (a), 5.6 (b), 5.7 (a) and 5.7 (b) are computed results with the proposed L_Q (.) function on growth and decay profiles. The results in [5.20] are also shown for comparison.

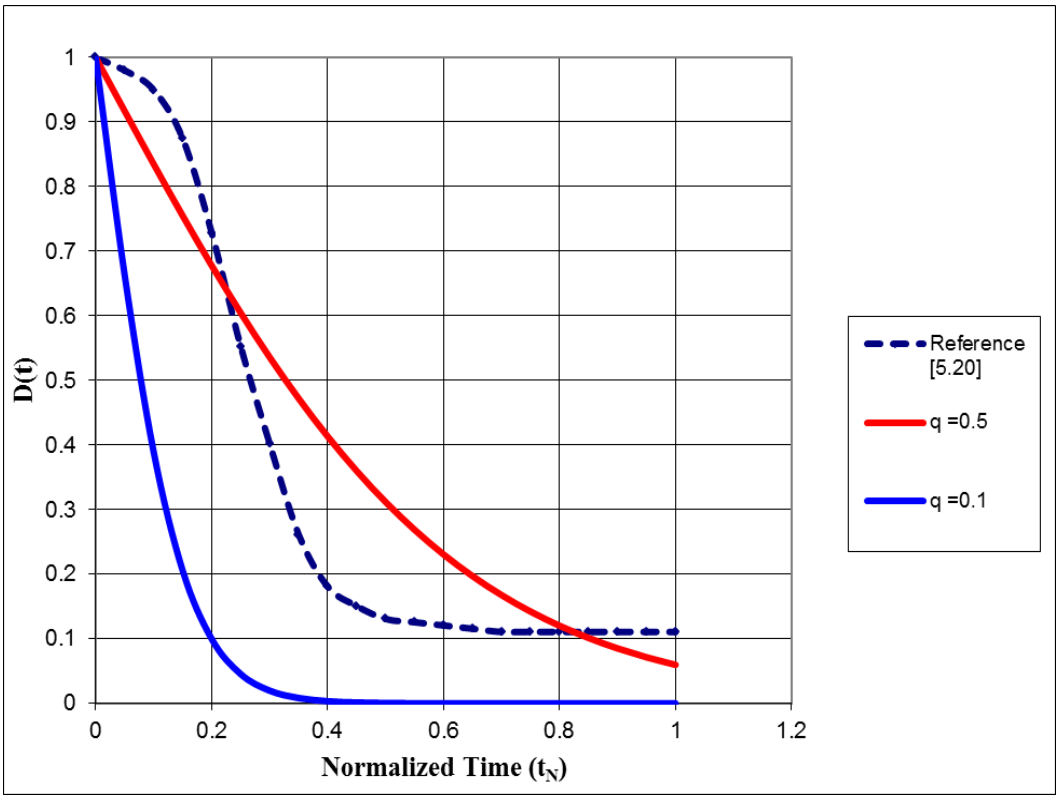

Figure 5.6 (b) Decay function curves

The results of [5.20] shown that the growth/decay dynamics falling in the regime outside the upper/lower bound constraint with Q = 0.5. For example, if Q = 0.1 is used, the results of [5.20] are closer to the proposed model with L_Q(.) function.

When Q → 0 means that the underlying statistics corresponds to non-interacting populations being present. For example in Figure 5.6 (a) considering the normalized time

(t_N), the results of the [5.20] are on the interaction domain around 0.2 to 0.3 and later, the dynamics corresponds to non- interaction regime. Similar observations can be made with growth/decay curves of Figures 5.6 (b), 5.7 (a) and 5.7 (b). It implies that in the initial stage the constituent of the cellular medium, namely infected and non-infected parts interact with the oncolytic viruses with a dispersion rate of b = 0.02 being present. After a certain growth/decay time, the interaction ceases.

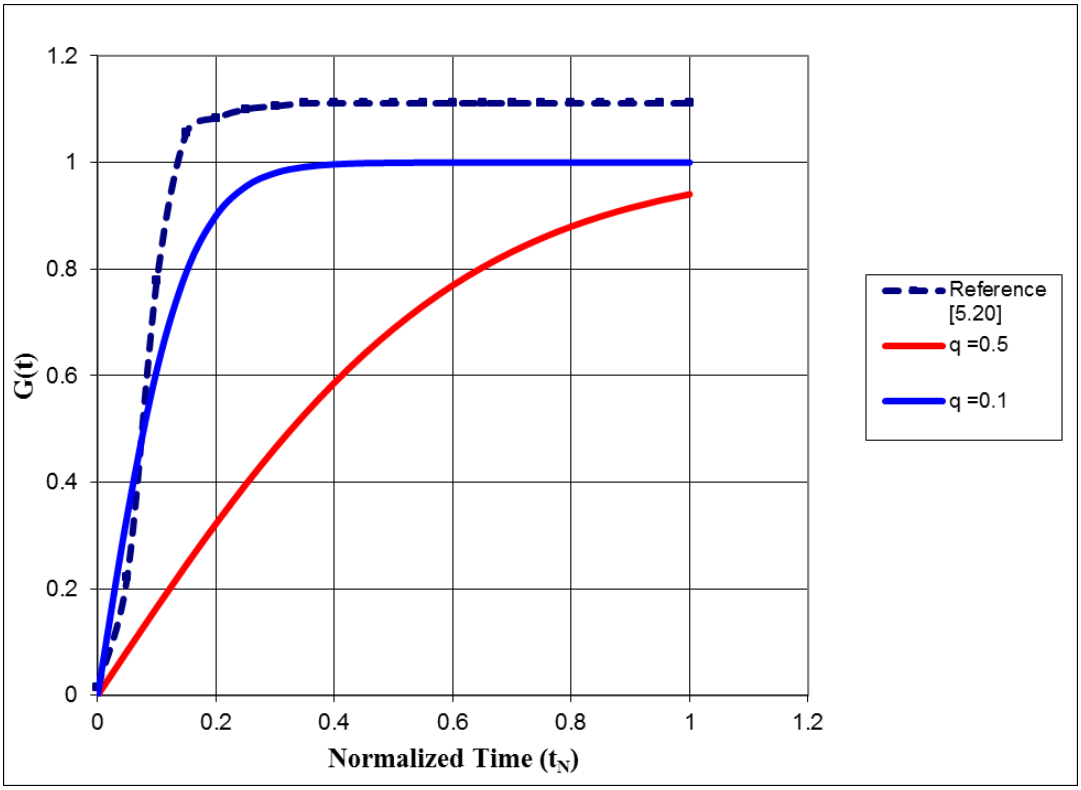

Figure 5.7 (a) Growth function curves

When the b-value, namely transmission rate of viral dispersions change to 0.06, the decay and growth curves are shown in Figure 5.7 (a) and in Figure 5.7 (b). In this case, the states of growth and decay conform to non-interaction of included constituents in the cells.

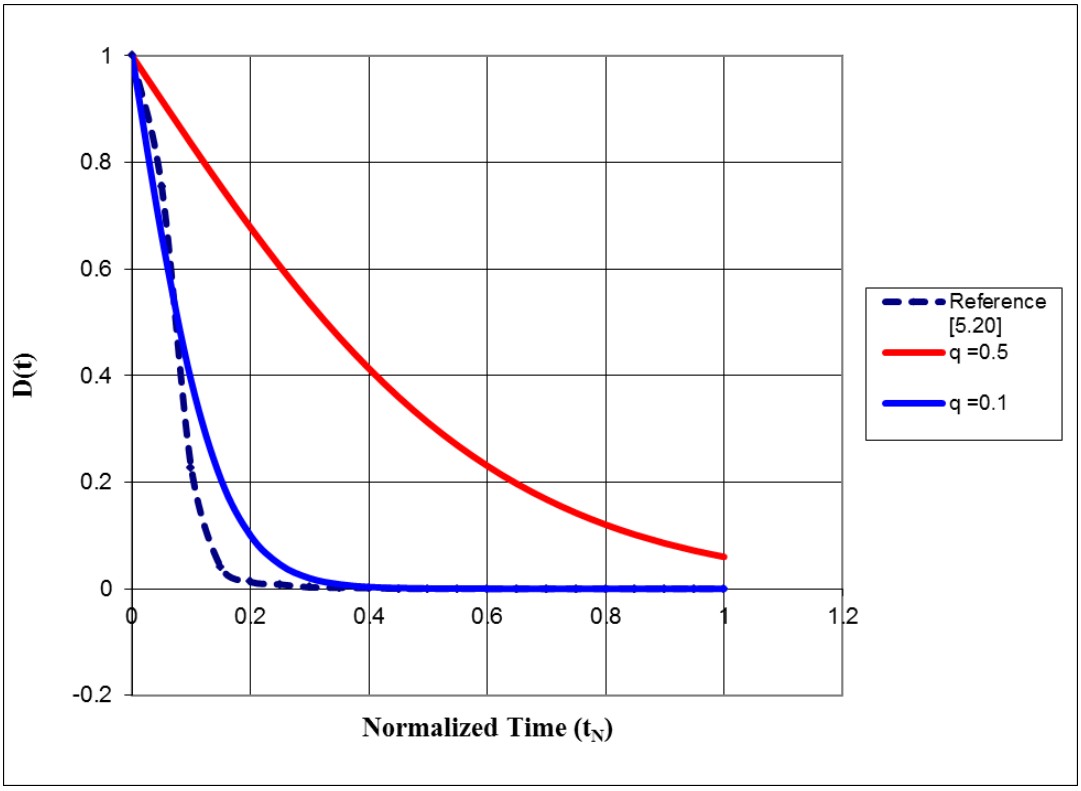

Figure 5.7 (b) Decay function curves

5.10 Discussions and Concluding Remarks

In summary presented in this chapter of the following research task and the results are following

- Statistical mixture theory is applied to the mixture constituents of a cytogenetic complex
- Both normal and contamination entities in the cytogenetic complex are considered. Hence, the presence of abnormal constituents such as cell aberrations, are estimated in terms of information redundancy parameter

- The mixture state of clonal and non-clonal alterations in cancer evolution is analyzed in the context of the cancer growth
- Specific to the temporal evolution of oncological state, calculations using available data are made with respect to deducing cancer growth or decay curve.

Such growth and decay curve are non-empirically indicated *via* stochastically justified logistic function namely the Langevin-Bernoulli function. Comparisons are made with available results.

CHAPTER VI

RESULTS, DISCUSSIONS AND CONCLUSIONS

6.1 Introduction

The studies performed in the earlier chapter are summarized and relevant deliberation of the outcome is discussed with inferential conclusions in this chapter. Further, open-questions on the research that provide a scope for further studies in this area are indicated.

Commensurate with the objectives, and the efforts presented in the previous chapters, the specific tasks carried out in this research can be enumerated as follows: Cytogenetics is a study on the genetic considerations associated with structural and functional aspects of the cells, specifically seeking out the so called normal and abnormal chromosomal features. Chromosomes refer to a gamut of structures in the cells that contain a vast extent of body's information stored in the form of strings of DNA. Within the scope of standard process of classifying each chromosome (known as *karyotyping*), normal chromosomes are traditionally identified and their structures are graphically represented.

Notwithstanding the existence of such normal karyotypes of a given species, chromosomal abnormality may prevail reflecting the presence of atypical number of chromosomes or structural abnormality in one or more chromosomes. Defined as *chromosomal aberrations* (CA), such anomaly may indicate certain genetic pathogenic conditions (known as *genetic disorders*), either inherited and/or induced as a result of

abiotic ambient such as ionizing radiations. Such disorders can be quantitatively specified in terms of the extent of chromosomal aberrations observed.

Further, presence of chromosomal aberrations implicitly depicts disorganization in the complexity of the underlying cytogenetics. The present study essentially addresses an analytical approach to quantify such disorganization *via* entropy considerations, using statistical mixture theory applied to the cellular contents of normal and aberrated chromosomes.

Both numerical anomaly of abnormal number of chromosomes (known as *aneuploidy*) as well as structural alterations seen in the physical form of chromosomes would imply a stochastical framework of biological complexity. That is, the associated entities constitute a set of cytogenetic factors of randomness implied as a result of aneuploidy and diverse structural features in the cellular complex. The coexistence of normal and abnormal chromosomes can be regarded as a state of statistical mixture.

Hence, informational entropy (in Shannon's sense) is prescribed to measure the extent of such chromosomal aberrations. That is, by comparing the molecular entropy at cytogenetic level of normal and anomalous chromosomes, one can specify/elucidate the intrinsic property of genetic disorders associated with the disorder in cell phenotype complex. The informational entropy is consistent with the principle of stochastical mixture theory.

Relevantly, the proposed approach assumes the presence of normal and abnormal chromosomal sets in varying proportions within the cytogenetic complex, and hence, the stochastical mixture theory is invoked to ascertain the information redundancy as a function of fractional abnormal chromosome population. Use of such computed details on

information redundancy of a cytogenetic system is indicated in this study as a track-parameter towards the progression of genetic disorder, for example, the macroscopically seen growth of cancer. Lastly, using the results obtained, conclusions are enumerated, inferences are outlined and directions for future studies are suggested.

6.2 Discussion and Inferences

The essence of the study is to indicate the relevance of bioinformatics in the cytogenetic framework. Such a framework of cytogenetic system can be considered as a complex system made of normal and aberrated cells and their interacting statistics.

In the conceived description of cytogenetic complex as above, the concept of informatics are introduced to ascertain the disorganized features in terms of entropy considerations or information contents in Shannon's sense.

Relevant details on the stochastical aspects of cellular complex, aberrated constituents (that may lead to genetic disorders etc.) and entropy features are used to model oncological growth.

Consistent with the above efforts the following can be stated as salient outcomes

- A comprehensive survey on cytogenetic complex viewed in the framework of bioinformatics
- Mining of data related to cytogenetic bioinformatics
- Compiling appropriate biostatistical methods for the modeling pursuits
- A simple proportion mixture model is developed to estimate the quantitative profiles of normal and abnormal chromosomes. Hence using the results in Nikitana et a.l. [4.1], the mathematical model concerning maternal cell contamination causing spontaneous abortions is revisited and the risk-factor is

specified within an upper and lower bound. Relevant outcomes can be considered as new

- The mixture model is made more comprehensive using statistical mixture theory and applied to cytogenetic context. Again considering clonal and non-clonal alterations relevant to cancer growth model is deduced using Langevin-Bernoulli function. More similar modeling is indicated on the results due to Di Pierdomenico et al. [5.9] and Agarwal et al. [5.20].

6.3 Open-questions for Future Study

Relevant to the study summarized above, the following open-questions can be indicated for future studies. The essence of the present research on the cytogenetic complex can be extended to study the associated bioinformatics relevant to the following

- Fuzzy aspects of overlaps of normal and abnormal contents and their functions in the cellular media
- Sudden epochal aberrations that may lead to acute conditions and genetic disorders: stability analysis
- Apart from oncogenetic viruses, the role of other virulent effects at cytogenetic level can be studied
- Systematic application of bioinformatics and biostatistics in the cytogenetic framework towards appropriate data-mining and bioinformatics.

6.4 Closure

The present study is a debut attempt in the fusing concepts of bioinformatics, biostatistics and data mining in the context of studying cytogenetics. Though not exhaustive, it offers a new avenue of thinking for further studies.

APPENDIX

COMPUTATION

Computations/results: Chapter 4: To calculate upper and lower bounds (UB and LB) of risk factors for spontaneous abortion. Consider the data for two groups of "46, XX" spontaneous abortions with or without Y chromosome(Y+ and Y−) are presented in table 4.1.

Case study 4.1 (a): Computation for group "46, XX" spontaneous abortions with Y chromosome (Y+).

Step I: Prepare an ensemble of 14 trails (x1, x2…x14) randomly selected within the given range. In case a range is not specified, consider the range to be ±2 from the given value. The risk factor mentioned below are exogenous characteristics(X_1 to X_9) represented in Table 4.1.

Risk Factor	x1	x2	x3	x4	x5	x6	x7
1	8.88	8.1	7.3	3.9	5.4	4.8	6.5
2	26	32	22	18	26	11	20
3	27.21	33	23	10	8	25	9
4	4	3.56	3.69	3.8	3.96	3.6	3.54
5	12	12.1	11.8	12.3	11.5	12.5	12.2
6	2	2.2	1.96	1.5	2.4	1.8	1.6
7	14	13.8	13.3	14.5	13.7	14.1	13.6
8	3	3.3	2.8	2.5	3.5	2.9	2.6
9	35.08	50	49	25	26	19	33

Risk Factor	x8	x9	x10	x11	x12	x13	x14
1	7.6	3.2	2.6	-1	1.5	-1	-2.6
2	9	14	6.5	25.5	25	5	-6.54
3	17	21	6.2	18	19	4	-6.36
4	3.2	3.7	3.42	3.48	3.81	3.95	3
5	11.6	12.4	11	11	10.8	11.2	10.5
6	2.3	1.7	2.4	1.8	1.6	1.9	1.5
7	14.3	14.2	14.4	12.8	13.5	12.6	12.5
8	3.8	2.7	3.2	2.4	1.8	2	1.5
9	29	31	22	1	5	-2	-16.8

Step II: Calculate upper and lower bound for all the 14 trails for both the cases using modified Bernoulli-Langevin function for determining the risk factor f(z)

$$f(z) = ½ + ½ L_q (z/2)$$

where, (z) = summation of the nine exogenous characteristics (X_1 to X_9) specified in Table 4.1. For upper bound (UB) consider q = ½ and for lower bound (LB) consider q = ∞.

Trial	Upper Bound q = ½	Lower Bound q = ∞
x1	0.844151804	0.582138002
x2	0.883997742	0.597549974
x3	0.848718121	0.583747958
x4	0.761640393	0.557348795
x5	0.781963539	0.562862814
x6	0.769036343	0.559321149
x7	0.785422283	0.56383232
x8	0.776055575	0.561228651
x9	0.789445602	0.564972509
x10	0.712640156	0.545088728
x11	0.721092961	0.547116873
x12	0.738823529	0.551480719
x13	0.616645523	0.523753673
x14	0.489346041	0.497864667

To calculate extrema (supremum and infimum) on the bounds UB and LB values:

Extrema refer to: Sup (*supremum*) and Inf (*infimum*) that correspond respectively to certain ± off-sets prescribed on the basis of statistical quantile on a given variable. Two quantiles are specified thereof:

- First level extremum corresponds to off-sets by 2nd quantile criterion with ± 47.73% off-sets
- Second level extremum corresponds to off-sets by 1st quantile criterion with ± 34.13% off-sets

Step I: To calculate the error margins for the UB by considering with ± 47.73% off-sets i.e. 2nd quantile [Z_{UB} (1 +0 .4773) and Z_{UB} (1 − 0.4773)].

- Similarly calculate the error margin for LB [Z_{LB} (1 + 0 .4773) and Z_{LB} (1 − 0.4773)].

Step II: Calculate the average of the higher error margin [Z_{UB} (1 + 0 .4773) and Z_{LB} (1 + 0.4773)] of both the bounds.

- Similarly calculate the error margin for LB [Z_{UB} (1 − 0 .4773) and Z_{LB} (1 − 0.4773)].

Step III: Repeat Step I and Step II by considering ± 34.13% off-sets i.e. 1st quantile.

Step VI: The final averages obtained from Step III are called final statistical limit i.e. supremum and infimum.

The calculated results for spontaneous abortions with Y chromosome (Y+) are shown below:

Trial	Final Statistical Limit: Infimum	Final Statistical Limit: Supremum
x1	0.693959529	0.499984116
x2	0.720845197	0.519354707
x3	0.696964587	0.5021492
x4	0.641752547	0.462370018
x5	0.654323587	0.471427204
x6	0.646310684	0.465654065
x7	0.656478147	0.472979521
x8	0.650653975	0.46878332
x9	0.658990446	0.474789583
x10	0.611946422	0.440895294
x11	0.617045916	0.444569378
x12	0.627795925	0.452314547
x13	0.554859809	0.399765519
x14	0.480326141	0.346065485

The results are shown in Figure 4.1(a) A risk-factor versus an ensemble of trials (T_1, T_2…, T_{14}) for (Y+) state.

Case study 4.1 (b): Computation for group "46, XX" spontaneous abortions without Y chromosome (Y−).

Risk factor	x1	x2	x3	x4	x5	x6	x7
1	8.86	5.54	3.82	6.89	4.62	2.18	6.33
2	25.54	11.98	6.21	20.94	13.52	23.45	15.32
3	28.28	16.32	23.41	24.61	7.53	18.74	21.78
4	20	19.62	20.34	20.11	19.87	20.63	19.54
5	62	61.44	61.29	62.35	62.18	62.64	62.41
6	12	12.34	12.48	11.79	11.43	11.68	11.92
7	64	64.21	64.65	63.82	63.52	63.38	64.38
8	30	29.11	30.26	29.48	29.45	30.34	30.12
9	24.05	16.82	18.54	11.67	22.43	18.23	14.67

Risk Factor	x8	x9	x10	x11	x12	x13	x14
1	3.01	7.21	2.76	7.9	5.9	-1	-2.76
2	18.22	9.63	5.27	19	1	-2	-5.27
3	10.24	13.65	5.28	2	4	7	-5.58
4	20.65	19.43	20.45	18.45	19	18.94	18
5	62.5	62.58	61.53	61.98	60.39	60.89	60
6	12.21	11.57	12.56	11	10.95	11.2	10.83
7	63.45	64.32	64.54	62.67	63.01	62.25	62
8	29.28	30.65	29.55	28.32	29	28.87	28
9	20.11	19.01	13.86	4.32	-6	-8.44	-13.86

Trial	Upper Bound $q = ½$	Lower Bound $q = \infty$
x1	0.974085863	0.662205515
x2	0.957131207	0.642464319
x3	0.959140217	0.644419841
x4	0.964562608	0.65012654
x5	0.955497837	0.640929415
x6	0.964376594	0.649919135
x7	0.962011776	0.647357863
x8	0.958412394	0.643702404
x9	0.957509757	0.642826919
x10	0.943160167	0.63062655
x11	0.943042723	0.630537652
x12	0.918437131	0.614512997
x13	0.908298141	0.609020766
x14	0.874578494	0.59359266

Trail	Final Statistical Limit: Infimum	Final Statistical Limit: Supremum
x1	0.796135533	0.573599905
x2	0.778281211	0.56073622
x3	0.78021015	0.562125982
x4	0.785624995	0.56602727
x5	0.776739692	0.559625586
x6	0.785433577	0.565889358
x7	0.783036795	0.564162524
x8	0.77950696	0.561619348
x9	0.778641817	0.560996029
x10	0.765723968	0.551688974
x11	0.765623572	0.551616641
x12	0.745854976	0.537373759
x13	0.738249625	0.531894254
x14	0.714336847	0.51466557

The results are shown in Figure 4.1(b) A risk-factor versus an ensemble of trials (T_1, T_2…, T_{14}) for (Y−) state.

Computations/results: Chapter 5: Case studies related to cancer growth models.

Cancer growth model I: To calculate growth and decay of cancer consisting of clonal and non-clonal constituents in complex system, using Bernoulli-Langevin function presented in Table 5.3.

Step I: Calculations related to normalized values for C, N and CH are represented as N*, C* and CH* in Table 5.2.

Step II: The terms τ, q, x, y, I, II, LQ(τ) and A(τ) are descried below:

$q = 2/3(N^* + C^* + CH^*)$

$x = (1+1/q)\ \tau$

$y = (1/q)\ \tau$

I = 1st term of $L_Q(x)$

II = 2nd term of $L_Q(x)$

$L_Q(\tau) = I - II$

$A(\tau) = I - L_Q(\tau)$

Present study cancer decay function: N:$1- L_Q(\tau)$

Present study cancer growth function: C: $L_Q(\tau)$

τ	q	x	y	I	II	$L_Q(t)$	A(t)	N: $1-L_Q(\tau)$	C: $L_Q(\tau)$
0.5	0.6666	1.25	0.75	2.947	2.361	0.5854	0.4145	0.9058	1.0794
1.5	0.6666	3.75	2.25	2.502	1.533	0.9690	0.0309	0.0676	1.7866
2.5	0.6666	6.25	3.75	2.500	1.501	0.9983	0.0016	0.0035	1.8406
3.5	0.6666	8.75	5.25	2.5	1.500	0.9999	8.25×10^{-5}	0.0009	1.8435
4.5	0.6666	11.5	6.75	2.5	1.500	0.9999	4.11×10^{-6}	8.99×10^{-6}	1.8436
5.5	0.6666	13.75	8.25	2.5	1.5	1	2.05×10^{-7}	4.47×10^{-7}	1.8437
6.5	0.6666	16.25	9.75	2.5	1.5	1	1.02×10^{-8}	2.23×10^{-8}	1.8437

The values obtained for cancer growth and decay are plotted in Figure 5.5.

Cancer growth model II: To calculate growth and decay function for cancer using Bernoulli-Langevin function.

Figure 5.6 (a) Growth function:

Normalized time	Reference [5.20]	q = 0.5	q = 0.1
0	0.013888889	0.000166669	0.0007
0.05	0.027777778	0.083153217	0.33760036
0.1	0.061111111	0.165236164	0.611075555
0.15	0.138888889	0.245229642	0.794589965
0.2	0.277777778	0.32221168	0.900311618
0.25	0.472222222	0.395394673	0.954604819
0.3	0.666666667	0.464150716	0.980270632
0.35	0.805555556	0.528024066	0.991712426
0.4	0.894444444	0.586731228	0.996605124
0.45	0.944444444	0.640150506	0.998635398
0.5	0.966666667	0.688303608	0.999459404
0.55	0.977777778	0.73133226	0.99978827
0.6	0.988888889	0.769472596	0.999917828
0.65	0.994444444	0.803029662	0.999968346
0.7	1	0.832353792	0.999987881
0.75	1	0.85781997	0.999995384
0.8	1	0.879810754	0.999998249
0.85	1	0.898702874	0.999999338
0.9	1	0.914857315	0.999999751
0.95	1	0.928612491	0.999999906
1	1	0.940280028	0.999999965

The values obtained for cancer growth function are plotted in Figure 5.6 (a).

Figure 5.6 (b) Decay function:

Normalized time	Reference [5.20]	q = 0.5	q = 0.1
0	1	0.999833331	0.9993
0.05	0.979899497	0.916846783	0.66239964
0.1	0.949748744	0.834763836	0.388924445
0.15	0.874371859	0.754770358	0.205410035
0.2	0.728643216	0.67778832	0.099688382
0.25	0.552763819	0.604605327	0.045395181
0.3	0.40201005	0.535849284	0.019729368
0.35	0.261306533	0.471975934	0.008287574
0.4	0.180904523	0.413268772	0.003394876
0.45	0.150753769	0.359849494	0.001364602
0.5	0.130653266	0.311696392	0.000540596
0.55	0.125628141	0.26866774	0.00021173
0.6	0.120603015	0.230527404	8.21717E-05
0.65	0.115577889	0.196970338	3.16544E-05
0.7	0.110552764	0.167646208	1.21194E-05
0.75	0.110552764	0.14218003	4.61642E-06
0.8	0.110552764	0.120189246	1.75085E-06
0.85	0.110552764	0.101297126	6.61602E-07
0.9	0.110552764	0.085142685	2.49215E-07
0.95	0.110552764	0.071387509	9.36199E-08
1	0.110552764	0.059719972	3.50862E-08

The values obtained for cancer decay function are plotted in Figure 5.6 (b).

Figure 5.7 (a) Growth function:

Normalized time	Reference [5.20]	q = 0.5	q = 0.1
0	0.013888889	0.000166669	0.0007
0.05	0.222222222	0.083153217	0.33760036
0.1	0.777777778	0.165236164	0.611075555
0.15	1.055555556	0.245229642	0.794589965
0.2	1.083333333	0.32221168	0.900311618
0.25	1.1	0.395394673	0.954604819
0.3	1.105555556	0.464150716	0.980270632
0.35	1.111111111	0.528024066	0.991712426
0.4	1.111111111	0.586731228	0.996605124
0.45	1.111111111	0.640150506	0.998635398
0.5	1.111111111	0.688303608	0.999459404
0.55	1.111111111	0.73133226	0.99978827
0.6	1.111111111	0.769472596	0.999917828
0.65	1.111111111	0.803029662	0.999968346
0.7	1.111111111	0.832353792	0.999987881
0.75	1.111111111	0.85781997	0.999995384
0.8	1.111111111	0.879810754	0.999998249
0.85	1.111111111	0.898702874	0.999999338
0.9	1.111111111	0.914857315	0.999999751
0.95	1.111111111	0.928612491	0.999999906
1	1.111111111	0.940280028	0.999999965

The values obtained for cancer growth function are plotted in Figure 5.7 (a).

Figure 5.7 (b) Decay function:

Normalized time	Reference [5.20]	q = 0.5	q = 0.1
0	1	0.999833331	0.9993
0.05	0.753768844	0.916846783	0.66239964
0.1	0.226130653	0.834763836	0.388924445
0.15	0.040201005	0.754770358	0.205410035
0.2	0.012562814	0.67778832	0.099688382
0.25	0.008040201	0.604605327	0.045395181
0.3	0.002512563	0.535849284	0.019729368
0.35	0.00201005	0.471975934	0.008287574
0.4	0.001507538	0.413268772	0.003394876
0.45	0	0.359849494	0.001364602
0.5	0	0.311696392	0.000540596
0.55	0	0.26866774	0.00021173
0.6	0	0.230527404	8.21717E-05
0.65	0	0.196970338	3.16544E-05
0.7	0	0.167646208	1.21194E-05
0.75	0	0.14218003	4.61642E-06
0.8	0	0.120189246	1.75085E-06
0.85	0	0.101297126	6.61602E-07
0.9	0	0.085142685	2.49215E-07
0.95	0	0.071387509	9.36199E-08
1	0	0.059719972	3.50862E-08

The values obtained for cancer decay function are plotted in Figure 5.7 (b).

REFERENCES

[1.1] A. T. Annunziato, DNA packaging: nucleosomes and chromatin, *Nature Education*, Vol. 1, 2008.

[1.2] A. J. F. Griffiths, W. M. Gelbart, J. H. Miller and R. C. Lewontin, *Modern Genetic Analysis*, W. H. Freeman and Company, New York, NY, 1999.

[2.1] C. O'Connor, Chromosomal segregation in mitosis: The role of centromeres, *Nature Education*, Vol. 1(1), 2008.

[2.2] C. Hoischen, M. Bussiek, A. Derome, F. Hayes and S. Diekmann, Bacterial centromeres and kinetochore complexes, Available at: *http://www.fli-leibniz.de/images/groups/diekmann/hoischen/ch_kinetochore.pdf*

[2.3] K. S. Bloom and J. Carbon, Yeast centromere DNA is in a unique and highly ordered structure in chromosomes and small circular mini chromosomes, *Cell*, Vol. 29(2), 1982, 305-317.

[2.4] *Mapping protein/DNA interactions by cross-linking, PMID: 21413366*, Institut national de la sante et de la recherche medicale (INSERM), Paris, 2001. Available at: http://www.ncbi.nlm.nih.gov/books/NBK7107/

[2.5] G, M. Copper, *The Cell: A Molecular Approach*, Sinauer Associates publishers, Sunderland, MA, 2000.

[2.6] A. F. Pluta, A. M. Mackay, A. M. Ainsztein, I. G. Goldberg and W. C. Earnshaw, The Centromere: hub of chromosomal activities, *Science*, Vol. 270, 1995, 1591-1594.

[2.7] J. C. Lamb, J. Theuri and J. A. Birchler, What's in a centromere? *Genome Biology*, Vol. 5 (9), 2004, 239-5.

[2.8] D. J. Amor and K. H. A. Choo, Neocentromeres: Role in human disease, evolution and centromere study, *Am J. Hum Genet.* Vol. 71, 2002, 695-714.

[2.9] S. K. Knutson, 2008, Available at: http://etd.library.vanderbilt.edu/available/etd-05222008-164304/unrestricted/

http://etd.library.vanderbilt.edu/available/etd-05222008-164304/unrestricted/01TitlePage.pdf

[2.10] T. Kouzarides, Chromatin modifications and their cell function, *Cell*, Vol.128, 2007, 693-705.

[2.11] L. C. Woodcock and R. P. Ghosh, Chromatin higher-order structure and dynamics, *Cold Spring Harb Perspect Biol*, 2010, 1-25.

[2.12] S. E.Torigoe, D. L. Urwin, H. Ishii, D. E. Smith and J. T. Kadonaga, Identification of a rapidly formed non-nucleosomal histone-DNA intermediate that is converted into chromatin by ACF, *Mol. Cell* Vol.43, 2011, 638-648.

[2.13] V. L. Chandler, Paramutation: From mazie to mice, *Cell*, Vol.128, 2007, 641-645.

[2.14] S. Sharma, T. K. Kelly and P. A. Jones, Epigenetics in cancer, *Oxford Journals*, Vol. 31 (1), 2009, 27-36.

[3.1] P. Stankiewicz and J. R. Lupski, Structural variation in the human genome and its role in disease, *Annual Review of Medicine*, Vol. 61, 2010, 437–455.

[3.2] G. H. Perry, N. J. Dominy, K. G. Claw, A. S. Lee, H. Fiegler, R. Redon, J. Werner, F. A. Villanea, J. L. Mountain, R. Misra, N. P. Carter, C. Lee and A. C. Stone, Diet and the evolution of human amylase gene copy number variation, *Nature Genetics*, Vol. 39, 2007, 1256-1260.

[3.3] A. R. Varsale, A. S. Wadnerkar and R. H. Mandage, Cancer investigation: A genome perspective, *Academic Journals*, Vol. 5(5), 2010, 79-86.

[3.4] R. Bishop, Applications of Fluorescence in situ hybridization (FISH) in detecting genetic aberrations of medical significance, *Oxford Journals*, Vol. 3(1), 2010, 85-95.

[3.5] H. H. Heng, J. B. Stevens, G. Liu, S. W. Bremer, K. J. Ye, P. V. Reddy, G. S. Wu, Y. A. Wang, M. A.Tainsky and C.J. Ye, Stochastical cancer progression driven by non-clonal chromosome aberrations, *J Cell Physiol*., Vol. 208 (2), 2006, 461-72.

[3.6] I. Lobo, Chromosome abnormalities and cancer cytogenetics, *Nature Education* Vol. 1(1), 2008.

[3.7] A. I. Baba and C. Catoi, *Comparative Oncology*, The Publishing House of the Romanian Academy, Bucharest, Romania, 2007.

[3.8] P. U. Devi, Basics of Carcinogenesis. *Health Administrator*, Vol. XVII (1), 2005, 16-24.

[3.9] Vogelstein B., Fearon E., Hamilton S., Kern S., Preisinger A., Leppert M., Nakamura Y., White R., Smits A., Bos J. Genetic alterations during colorectal-tumor development. *New Engl. J. Med*, Vol. 319 (9), 1988, 525-532.

[3.10] F. Jiang, R. Desper, C. H. Papadimitriou, A. A. Schaffer, O. P. Kallioniemi, J. Richter, P. Schraml, G. Sauter, M. J. Mihatsch and H. Moch, Construction of evolutionary tree models for renal cell carcinoma from comparative genomic hybridization data, *Cancer Research*, Vol. 60 (22),2000, 6503-9.

[4.1] T. V. Nikitana, I. N. Lebedev, N. N. Sukhanova, E. A. Sazhenova, and S. A. Nazarenko, A mathematical model for evaluation of maternal cell contamination in cultured cells from spontaneous abortions: significance for cytogenetic analysis of prenatal selection factors, *Fertility and Sterility*, Vol. 83(4), 2005, 964-972.

[4.2] A. N. Dupont, Risk-Evaluation in Clinical Diagnostic Studies: Ascertaining Statistical Bounds via Logistic Regression of Medical Informatics Data, 2011; 96, 1507508.

[5.1] K. Lichtenecker and K. Rother, Die Herleitung des Logarithmischen Mischungsgesetzes aus allegemeinen Prinzipien der stationären Strömung, *Physik Zeitschrift*,, Vol. 32, 1938, 255-260.

[5.2a] P. S. Neelakanta, Complex permittivity of a conductor-loaded dielectric, *Journal of Physics*, Vol. 2, 1990, pages 4935-4947.

[5.2b] P. S. Neelakanta, R.I Turkman and T. K Sarkar, Complex permittivity of a dielectric-mixture correlated version of logarithmic law of mixing, *Electronic Letters*, Vol. 21(7), 1985, 270-271.

[5.3] P. S. Neelakanta, T. V. Arredondo and D. DeGroff, Redundancy attributes of a complex system: Application to bioinformatics, *Complex Systems*, Vol. 14, 2003, 215–233.

[5.4] P. S. Neelakanta, Application of logarithmic law of mixing to electric susceptibility. *Electronics Letters*, Vol. 25, 1989, 800-802.

[5.5] P.S. Neelakanta, "Permittivity of Dielectric-conductor Mixture: Application of Logarithmic Law of Mixing to Electrical Susceptibility," *Electronics Letters*, Vol. 25, 1989, 800-8022.

[5.6] P. S. Neelakanta, J. N. Kapur, and H. K. Kesavan, *Entropy Optimization Principles with Applications* (Academic press/Harcourt Brace Jovanovich publishers, Boston, MA: 1992).

[5.7] B. S. Taylor, Jordi, Barretina, N. D. Socci, P. DeCarolis, M. Ladanyl, M. Meyerson, S. Singer and C. Sander, Functional copy-number alterations in cancer *EMBS*, Vol. 3, 2008, 3179.

[5.8] P. V. Loo, S. H. Nordgard, O. C. Lingjaerde, H. G. Russnes, I. H. Rye, W. Sun, V. J. Weigman, P. Marynen, A. Zetterberg, B. Naume, C. M. Perou, A. B. Dale and V. N. Kristensen, Allele-specific copy number analysis of tumors, *PNAS*, Vol.107, 2010, 39.

[5.9] J. DiPierdomenico, H. Ying, F. Lin, H. H. Q. Heng, A mathematical model relating chromosome aberrations to cancer progression, *Proceedings of the 28th IEEE*, 1-4244-0033-3, 2006, 2028-2031.

[5.10] A. E. Ferdinand, "The theory of system complexity, *International Journal of General Systems,* Vol.1, 1994, 19-235.

[5.11] P. S. Neelakanta, *Information Theoretic Aspects of Neural Networks*, (CRC Press, Boca Raton, FL, 1999.

[5.12] R. M. Bendett and P. S. Neelakanta, A relative complexity metric for decision-theoretic applications in complex systems, *Complex Systems*, Vol.12, 2000, 281-295.

[5.13] E. Jaynes, "On the rationale of maximum entropy methods," *Proceedings of IEEE*, Vol. 70, 1982, 939-952.

[5.14] T. M. Cover and J. Thomas, *Elements of Information Theory*, John Wiley and Sons Inc., New York, NY, 1991.

[5.15] S. Kullback, *Information Theory and Statistics,* Wiley Interscience Publications, NewYork, NY, 1959.

[5.16] J. Lin, "Divergence measure based on the Shannon entropy," *IEEE Transactions on Information Theory*, Vol. 37, 1991, 145-151.

[5.17] L. Xiao.bo, Mathematical modeling of carcinogenesis based on chromosome aberration data, *The Chinese Journal of Cancer Research,* Vol. 21(3), 2009, 240-246.

[5.18] B. Vogelstein, E. R. Feason, S. R. Hamilton, etal. Genetic alterations during colorectal-tumor development, *New England Journal of Medicine,* Vol. 319(9), 1988, 525-32.

[5.19] M. Hoglund, A. Frigyesi, T. Sall, D. Gisselsson and F. Mitelman, Statistical behavior of complex cancer karyotypes. *Journal Genes, Chromosomes and Cancer,* Vol. 42, 2005; 327-341.

[5.20] M. Agarwal, A. S. Bhadauria, Mathematical modeling and analysis of tumor therapy with oncolytic virus, *Journal Applied Mathematics,* Vol. 2, 2011, 131-140.

[5.21] A. R. A. Anderson, and M. A. J. Chaplin, Continuous and discrete mathematical models of tumour- induced angiogenesis, *Bulletin. Mathematical Biology*, Vol. 60, 1998, 857-900.

[5.22] H. G. Othmer, and A. Stevens, Aggregation blowup, and collapse: The ABC's of taxis in reinforced random walks, *SIAM Journal of Applied. Mathematics* Vol. 57 (4), 1997, 1044-1081.

[5.23] A. Kubo, Mathematical analysis of some models of tumour growth and simultations, Tumour angiogenesis dynamics, *WSEAJ transactions on BIOLOGY and BIOMEDICINE*, Vol. 7(2), 2010, 31-40.

[5.24] A. A Patel, E. T. Gawlinskis, S. K. Lemieux and R. A. Gaten, A Cellular Automaton Model of early tumor growth and invasion; The effects of native tissue vascularity and increased Anaerobic tumor metabolism, *Journal of Theoretical. Biology* Vol. 213(3), 2001, 315-331.

[5.25] M. A. A. Castro, T. G. H. Onsten, J. C. F. Moreira and R. M. C. de Almeida, Chromosomal aberrations in solid tumors have a stochastical nature, *Mutation Research/ Fundamental and Molecular Mechanisms of Mutagenesis,* Vol. 600 (1-2), 2006, pages 150-164.

[5.26] F. Mitelman, Recurrent chromosome aberrations in cancer, *Mutation Research/ Reviews in mutation research*, Vol. 462 (2-3), 2000, 247-253.

[5.27] N. Fukshansky and W. Bar, DNA Mixtures: biostatistics for mixed stains with haplotypic genetic makers. *International Journal of Legal Medicine*, Vol. 119, 2005, 285-290.

[5.28a] P. S. Neelakanta and R. Yassin, A co-evolution model of competitive mobile platforms: Technoeconomic perspective, *Journal of Theoretical and Application electronic Commerce research*, Vol. 6(2), 2011, 31-49.

[5.28b] P. S. Neelakanta and R. C. T. Sardenberg, Consumer benefit versus price elasticity of demand: a nonlinear complex system model of pricing Internet services on QoS-centric architecture, *Netnomics*, Vol. 2(1), 2011, 31-60.

[5.29] P. S. Neelakanta and D. F. De Groff, *Neural Network Modeling*: *Statistical Mechanics and Cybernetics Perspectives*, CRC Press, Boca Raton, FL, 1994.