Spherical Harmonics for 3D Modeling of Chromosomes Territories in Normal and Aneuploid Nuclei

A Dissertation

Presented to

the Faculty of the Department of Biomedical Engineering

University of Houston

In Partial Fulfillment

of the Requirement for the Degree

Doctor of Philosophy

in Biomedical Engineering

by

Fan-yun Yen

May 2019

Spherical Harmonics for 3D Modeling of Chromosomes Territories in Normal and Aneuploid Nuclei

_____

Fan-yun Yen

Approved:

_____
Chair of the Committee
Fatima Merchant, Associate Professor,
Department of Engineering Technology
Department of Biomedical Engineering

Committee Members:

_____
Debananda Pati, Professor,
Department of Pediatric Hematology-Oncology,
Baylor College of Medicine

_____
Elebeoba E. May, Associate Professor,
Department of Biomedical Engineering

_____
Yingchun Zhang, Associate Professor,
Department of Biomedical Engineering

_____
Guoning Chen, Associate Professor,
Department of Computer Science

_____
Suresh K. Khator, Associate Dean,
Cullen College of Engineering

_____
Metin Akay, Professor and
John S Dunn Endowed Chair,
Department of Biomedical Engineering

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor Dr. Merchant for her continuous support of my Ph.D study and related research, and her patience, motivation, and immense knowledge. Her guidance helped me throughout my research and writing of this dissertation. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr. Pati, Dr. May, Dr. Zhang, and Dr. Chen for their insightful comments and encouragement. Dr. Pati's support, especially related to the pathophysiology of aneuploidy and providing the cell sample was critical to moving the computational model to practical application. Dr. Chen introduced me to computational visualization and the topic of spherical harmonics. Without their help and support, I could not complete this research.

My sincere thanks also go to my advisor during my undergraduate years at Ming Chung University, Dr. Tang who provided me with the opportunity to join his lab and start conducting research. He encouraged me to consider higher studies abroad and supported me at my senior year. I also thank my fellow lab mates for the productive discussions, the insightful suggestions, and their companionship in the lab over the last five years.

A special thanks to the Lee family including Tracy, Jason, Helen, and Elbert for their continuous support. They've been supporting me since the first day I arrived in the US. I wouldn't have been able to complete this achievement without their assistance. I also thank the Chen family including Michael, Tiffany, May, Super, and Hyper for their companionship and support throughout the years.

I thank my significant other, Emily for her continuous support and understanding. She has always been by my side. She showed me the beauty of Houston and brought happiness to my life.

In the end, I am grateful to my family for their selfless support both mentally and physically. They have made me who I am today. I consider myself nothing without them. I thank my brother for inspiring conversations, honest suggestions, and sharing invaluable tips on life. I thank my parents for bringing me into this world and their moral support, encouragement, and motivation. I thank all my family and friends. Thank you.

Spherical Harmonics for 3D Modeling of Chromosomes Territories in Normal and Aneuploid Nuclei

An Abstract

of a

Dissertation

Presented to

the Faculty of the Department of Biomedical Engineering

University of Houston

In Partial Fulfillment

of the Requirement for the Degree

Doctor of Philosophy

in Biomedical Engineering

by

Fan-yun Yen

May 2019

Abstract

DNA is packaged into chromosomes, which occupy a specific region in the three-dimensional (3D) nuclear space known as the chromosome territories (CTs). The spatial organization (SO) of CTs within the nucleus is non-random and any disruption of this organization leads to undesired changes, such as disease states. Determining how CTs organize in the nucleus can allow us to unravel any changes occurring during aneuploidy (loss or gain of chromosomes), a hallmark of cancer. Here, we describe a 3D modeling approach to allow precise shape estimation and localization of CTs in the nucleus of human embryonic stem cells (hES) undergoing progressive but defined aneuploidy. The hES cell line WA09 acquires an extra copy of chromosome 12 in culture with increasing passages. Both diploid and aneuploid nuclei were analyzed to quantitate the differences in the localization of CTs for chromosome 12 as it transitions from euploidy to aneuploidy. The CTs were detected with chromosome specific DNA probes via multi-color fluorescence in situ hybridization (FISH) in conjunction with confocal microscopy. We employed spherical harmonic (SPHARM) surface modeling to generate a well-defined 3D surface for both the nuclei and enclosed CTs, thereby allowing precise quantification of their size and shape. The estimated models were compared across multiple cells by aligning the nuclei to a well-defined template followed by determining CT position with respect to a local landmark. Our results present evidence of statistically significant changes in the spatial organization of CTs in trisomy-12 cells when compared to diploid cells from the same population. Additionally we observed, that changes in CT proximity relationships may affect gene expression of co-regulated genes.

Table of Contents

List of Figures

List of Tables

Chapter 1 – Introduction

The entire human genome that consist of about 3.5 meters long DNA has to be accommodated in the cell's nucleus within a small space that is one-hundredth of a millimeter in diameter. The DNA is packed into chromosomes that are distributed as chromosome territories (CTs) within the three-dimensional (3D) nuclear space [1]. The idea of compartmentalization of chromosomes was first introduced in 1885 and the term "chromosome territory" was formally defined in 1909 with a study of blastomere stages of the horse roundworm *Parascaris equorum*. However, due to the lack of experimental evidence, the concept of chromosome territory was almost been disapproved during the mid-twenties. New experimental support revived the study of chromosome organization. In 1980, the development of fluorescence in situ hybridization (FISH) enabled the direct detection and visualization of specific DNA sequences and later three-dimensional (3D) FISH allowed further investigations of chromosome arrangement in the 3D space within nucleus.

Current studies have suggested that the spatial organization of chromosome territories is non-random and can influence gene regulation, expression, and genomic instability. Studies investigating the radial positioning of CTs of specific cell types within nucleus have demonstrated correlations between CT position and chromosome gene content[2], [3] and chromosome size[4]–[7]. Moreover changes in the radial positioning of chromosomes is also seen across normal and malignant cells[8], and the radial distance of CTs has been reported as a factor influencing chromosome translocations [9]. Also, proximity patterns describing neighborhood arrangements of CTs have been established [10]–[12]. An important finding is that CTs have been thought to reposition in cancer and other disease

states. Although alterations in gene expression mostly ensue in aneuploidy, limited work is reported in terms of the correlation between CT positioning and aneuploidy. Therefore, further studies in both normal and abnormal cells to determine the relationship between CT positioning and genome abnormalities are clearly needed.

In this study, we utilize a human embryonic stem (hES) cell model, wherein aneuploid cells spontaneously and progressively acquire trisomy over extended culture to investigate the changes in the spatial organization of CTs. In order to analyze how aneuploidy affects the spatial organization of CTs, we introduce a 3D modeling framework that provides precise shape estimation based on the spherical harmonics shape descriptor of the 3D surface of CTs and the bounding nuclear surface. A key feature of the presented computational framework is that it allows a comprehensive comparison of CT spatial organization across multiple nuclei from a given population and/or between populations. Another feature of the framework is that the relative position of CTs is determined in a specific (x, y, z) coordinates. Finally, CT positioning data is analyzed relative to a set of microarray data that shows the gene expression of hES cells through a time series (passages), to investigate any influence of CT positioning with gene expression.

Chapter 2 – Background

2.1 Human Genome

The definition of genome given by the Medical Subject Heading vocabulary established by the National Library of Medcine is "the genetic complement of an organism, including all of its genes." A gene is a nucleic acid sequence in a deoxyribonucleic acid (DNA) molecule that carries certain hereditary information of the organism. DNA nucleotides come in 4 different types: adenine (A), cytosine (C), thymine (T), and guanine (G). Two nucleotides on opposite complementary DNA strands are connected with hydrogen bonds and each pair of the two nucleotides that connected is called a base pair (bp). The connecting rule for a base pair is: nucleotide type A and T can form a base pair and type C is connected to type G. The DNA thread in form of double helix wraps around structural proteins called histones that fold and pack the DNA into a structure known as chromatin. For most of the time in a cell cycle, the complex of DNA strand and histones exist in the nucleus in form of chromatin, which is loosely packed. The chromatin becomes highly compacted when the cell is undergoing nuclear division (mitosis) and is called chromosome.

Human genome consists of approximately $3.2\ x\ 10^9$ nucleotides that is distributed across 24 different chromosomes. Each somatic human cell contains two copies of each chromosome, whereas germ cells contain a single copy. Figure 2.1[13] shows a karyotype of the human genome. out of the 24 different chromosomes in human cells, 22 of them are common to both male and females and always exist in pairs, and the remaining two are the sex chromosomes, chromosome X and chromosome Y. Females carry two copies of chromosome X and males have one chromosome X and one chromosome Y in the nucleus.

The 22 common chromosomes (23 for female counting sex chromosome) are called homologous chromosomes, and the only non-homologous chromosomes are the sex chromosome in male.

One important role of chromosomes is to carry a set of genes that are the functional units of heredity. A gene is a segment of DNA sequence that has the instruction for producing a particular protein (for most cases) or, in some cases, RNA as the final product.

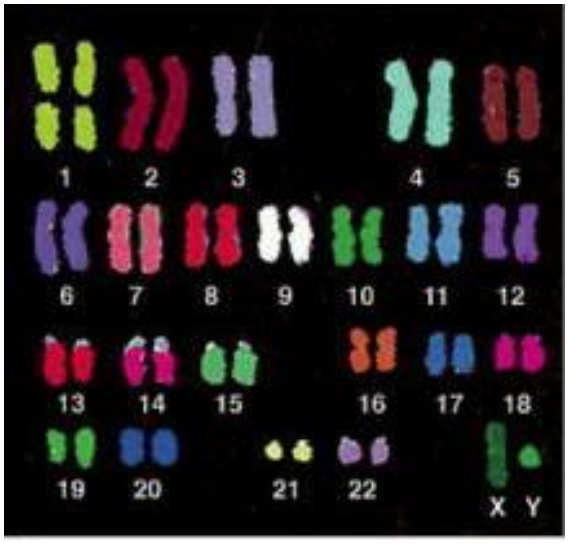

Figure 2.1. Full human chromosomes that are line up in name and color (figure from [13]).

2.2 Cell Cycle

During cell division genetic material (chromosomes) are passed on to the daughter cells. The life cycle of a cell involves series of stages of growth and development of the cell between the times when it is created to the moment it reproduces. In general, during the cell cycle, a cell has to grow, duplicate its genetic material, and split into two daughter cells. The term “cycle” conveys the idea that this whole process is not a linear process that stops, instead, the daughter cells repeat these series of action from the beginning of the cycle over again.

In eukaryotic cells, the cell cycle can be divided into two major phases: interphase and the mitotic phase (M phase). During interphase, the cells grow, replicate the DNA, and prepare for the mitotic (dividing) phase. The genes inside the cell actively express and synthesize proteins. Cell division occurs during the mitotic phase. Chromosomes become highly condensed then the cell physically splits into two daughter cells.

2.3 Chromosome Territories

In contrast to the M phase, chromosomes are in a less condensed form during interphase. These less dense forms of chromosomes occupy specific space within the nucleus and are arranged in a non-random fashion[5]. The region within the nucleus that each chromosome occupies is known as "chromosome territory [1]."

The idea of territorial arrangement of chromosome during the interphase of animal cells came from Carl Rabl in 1885 and the term "chromosome territory (CT)" was introduced by Theodor Boveri in 1909. During the 1970s and 1980s, most of electron microscopic evidence argued the chromosomes exist in nucleus in fiber-like form with no individual appearance during interphase, which refers to the "spaghetti style." Figure 2.2 uses wool thread to demonstrate the idea of territorial and non-territorial chromosome arrangement models. Since very limited evidence supported the concept of CT during that time, the idea of chromosome territories almost died out. In late 1970s and early 1980s, Stack and Cremer provided experimental evidence to support the concept of chromosome territories with different approaches[14]–[16].

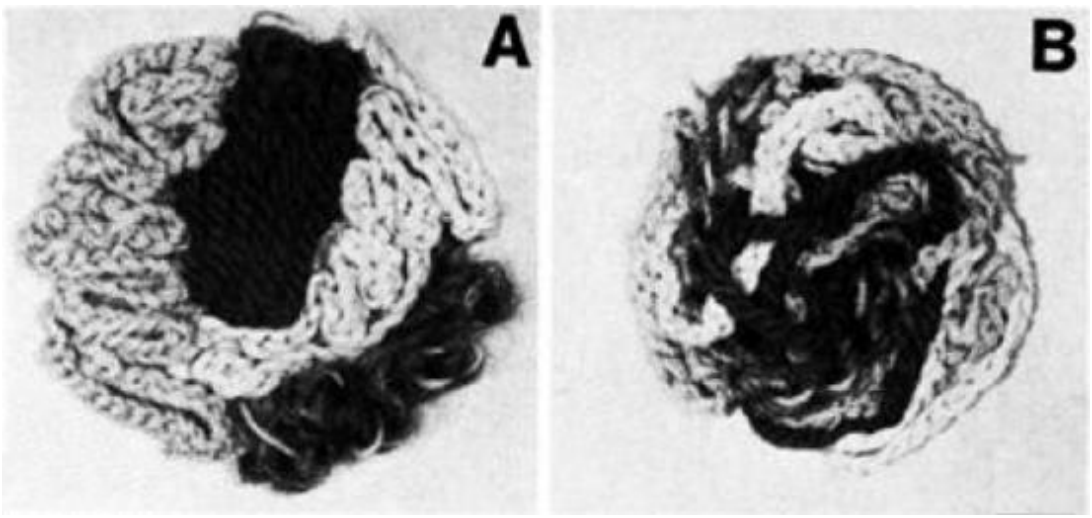

Figure 2.2. Demonstration of territorial and non-territorial chromosome organization in cell nuclei: (A) chromosome territory model and (B) spaghetti model (figure from [17]).

With the development of in suit hybridization, individual chromosomes were directly visualized. The fluorescence in situ hybridization (FISH) technique was then being expanded to three-dimensional (3D) aspect by combining with confocal laser microscopy to obtain series of optical sections[7]. 3D-FISH is now the most widely used technique for the study of CT organization during interphase.

2.4 Fluorescence in Situ Hybridization

One of the prevailing methods for detecting chromosomal abnormalities such as an increase or decrease in the number of chromosomes is using chromosome specific DNA probes via multi-color fluorescence in situ hybridization. In FISH, a DNA probe, that contains complementary sequence to the target chromosome, is designed to recognize and hybridize with the target chromosome within the interphase nucleus. The location of the target chromosome can be visualized directly or indirectly through fluorescence immunocytochemical techniques.

2.5 Non-random Organization of Chromosome Territories

To study the arrangement of chromosome within nuclei, it is important to determine a single or multiple reference points. This could be the gravity center of FISH signal or its surface. The non-randomness of CT organization has been determined in terms of (1) CT radial arrangement and (2) the proximity patterns. Non-random CT radial arrangement has been observed in correlation to gene density[3], [18], and various parameters such as transcriptional activity and replication timing[19]–[21]. On the other hand, the evidence that supports non-random proximity patterns is relatively less reported on small subsets. Notably, most proximity association were determined based on probabilistic methods from a population of nuclei, rather than a continuous observation on each nuclei. Also, the internal structure of chromosome and the idea of chromatin domains, considered as the second order structure of chromosome organization, has been studied has by many researchers. Chromosome domains are distinct form of each chromosome known as topologically associating domains (TADs)[22]–[24]. The non-randomness of the 3D arrangement in chromatin domain clusters also has recently proven by Cremer, 2019[25].

Researchers have found that the special organization (SO) of individual CTs is arranged in a non-random fashion and disruption of the arrangement of CTs may impact gene-expression patterns[26] and DNA stability[27]. Non-randomly distributed CTs within the 3D-nuclear space are observed in mammalian, invertebrate, yeast, and plant cells[1], [5], [28], [29], and specific to both cell [28], [30], [31] and tissue types[32] during interphase. Evidence shows that patterns can be recognized in the arrangement of CTs[5] in correlation to gene density[2], [33], chromosome size[4] and number[34]. Prior research also generally confirms the effects of SO on gene regulation [12], [35], [36]. A key finding is that

disruption in the SO results in unwanted genomic changes as seen in disease states like cancer [37].

2.6 Related Work

Many researches have studied the spatial organization of chromosome territories within nuclear space with different approaches. The most popular ones are (1) studying the radial distribution of CT within the nucleus, (2) finding the CT location with respect to some nuclear landmarks (e.g., envelop and nucleolus), (3) investigating the CT proximity, and (4) establishing probability map of CT organization.

Previous studies have investigated in the characteristics of positioning of genomic components in terms of radial positioning that derives from a ratio between the center and border of the nucleus[6], [8]. For example, Cremer *et al.* presented quantitative assessment to evaluate the 3D radial distributions of painted chromosome territories by a voxel-based algorithm. The nuclear space was divided into 25 equidistance shells with a thickness of Δr as a fraction of radius from the center of the nucleus to its border. The radial position of CTs were evaluated according to the shell that they were located within, and the gene content of the CTs was defined by the sum of the FISH signal intensities of all voxels[5]. Figure 2.3 shows the scheme of how the radial distribution of chromosome territories were defined. Each voxel position is defined by the relative distance to the radius, $r = r_1/r_0 * 100$, where r is the CT voxel position, r_0 defined by the nuclear radius, and r_1 represents the distance from the nuclear centroid to the CT voxel. DNA content of each shell was the integration of voxel intensities within one equidistance shell.

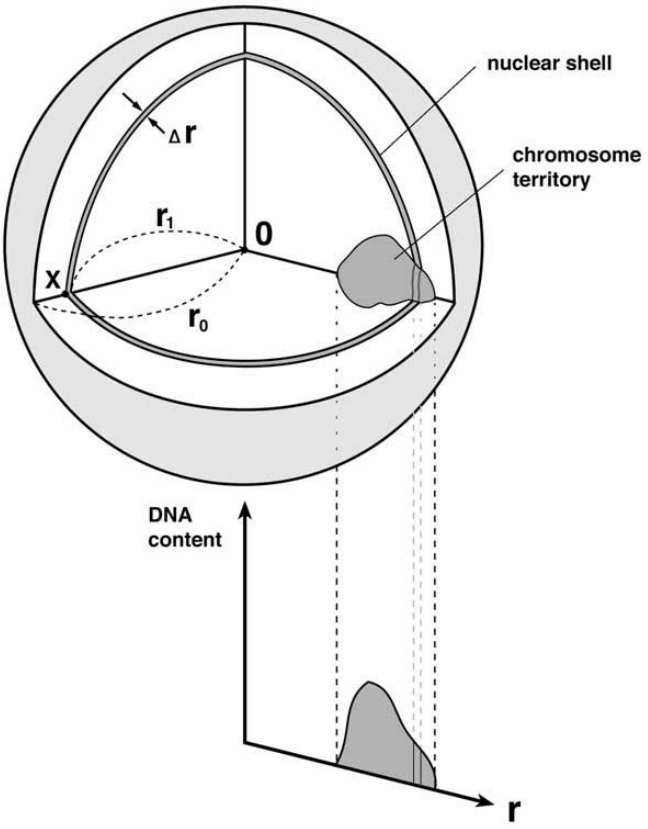

Figure 2.3. Scheme of 3D evaluation of the chromosome territories arrangements in spherical nuclei [5].

Another popular approach is studying the relative positioning with respect to other chromosomes [38] or to landmarks such as the nuclear envelope or nucleolus [39]. Brianna *et al*. segmented the FISH image using the Wavelet Transform Modulus Maxima algorithm [40] and investigated the CT neighborhood composition by categorizing CT proximity behavior according to the relative distance between CTs. As shown in Figure 2.4, the distance relationship between two chromosomes can be categorized to dispersed, homologous, heterologous, and clustered. Figure 2.4 (A) shows sample images of nuclei stained for specific chromosome pairs and an example of observed categorized CT patterns. Figure 2.4 (B) is the diagrammatic representation of CT proximity pattern categories. And Figure 2.4 (C) shows the frequency of patterns found in each group of a pair of

chromosomes. The resulting frequency of each category was coupled with gene translocation data to evaluate the relationship of CT proximity to chromosomal translocation.

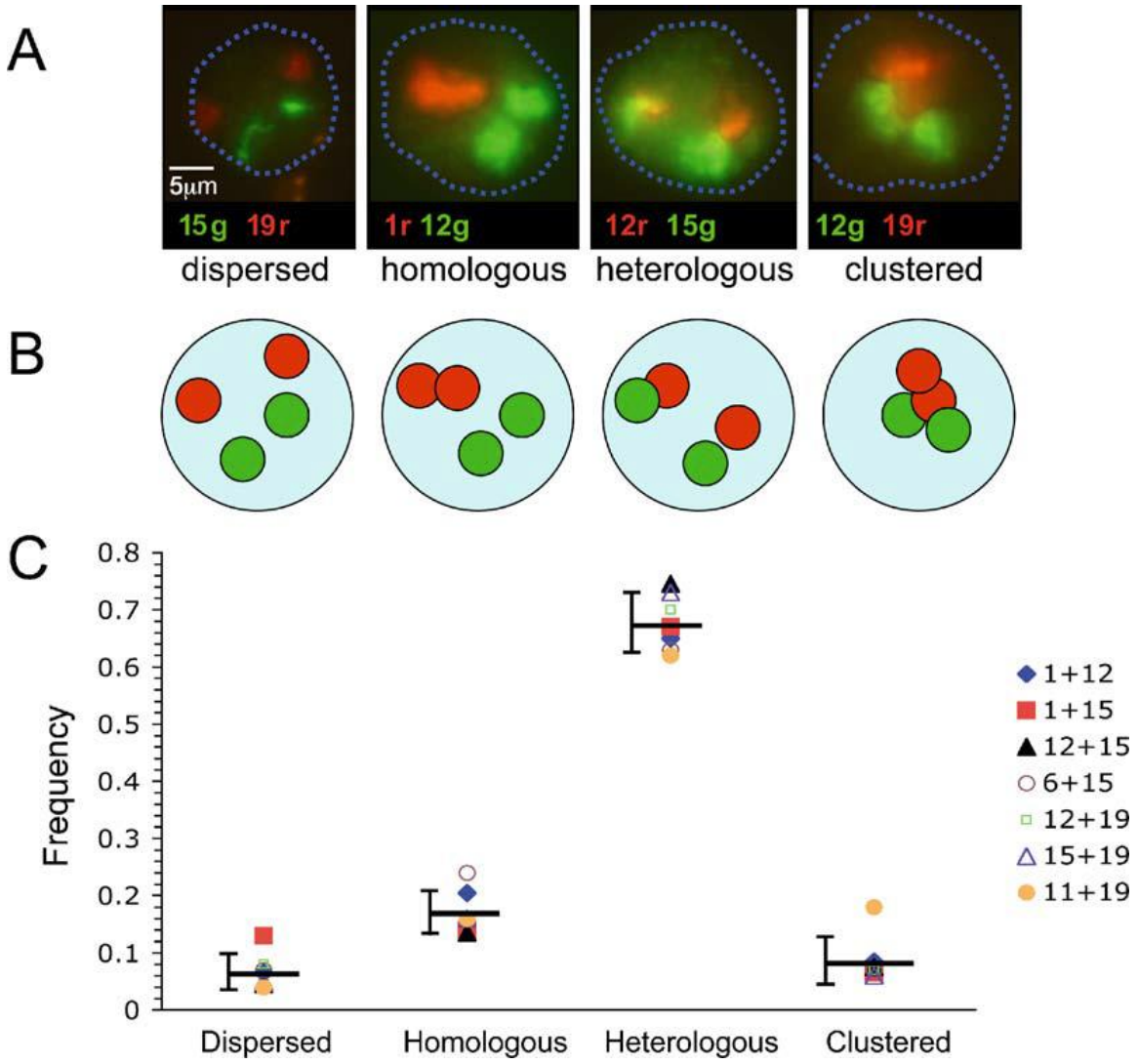

Figure 2.4. CT categorization based on CT proximity behavior [38].

Berger *et al.* utilized nucleolus as the nuclear landmark and aligned all the nuclear samples to the same direction at the half axis as shown in Figure 2.5(a). The central axis (dashed line) is defined by the line that links the nuclear center (green X) and the centroid (red X) of the nucleolus (shaded, red). Due to lack of a third nuclear landmark, the angle

around central axis is undefined (angle φ shown in Figure 2.5(a)). The 3D position of the genes shown in Figure 2.5 (b) were projected to a half circle according to the magnitude (R) and angle (α) of each gene 3D position as seen in Figure 2.5(c) and a probability density map (Figure 2.5 (d)) was created by constructing the isocontours of a 2D histogram based on the sum of Gaussian spots centered at the projection point location for given samples.

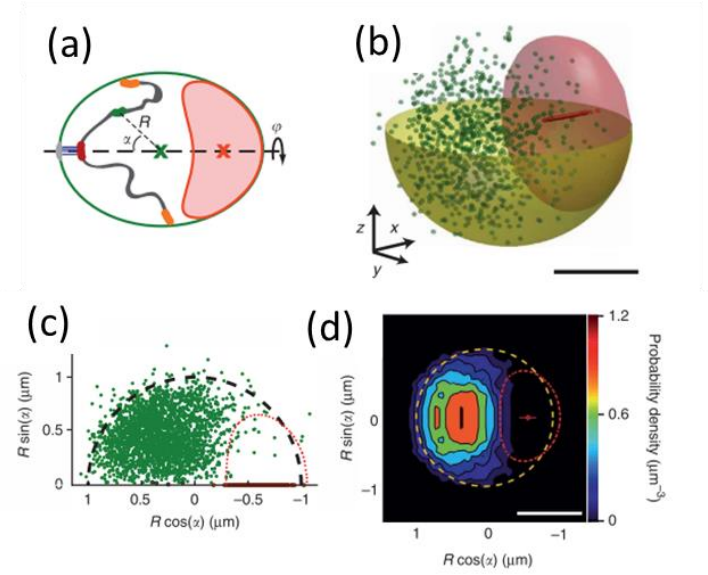

Figure 2.5. Nuclear landmark alignment and mapping by Brianna *et al*[38].

Due to the lack of incontrovertible structural landmarks in nuclei, an absolute genomic region is impossible to be defined, and, thus, most studies present probability distribution maps for the spatial organization of genomes [41], [42]. Zeitz *et al*. presented a probabilistic model for the arrangement of CTs with a geometric graph based algorithm termed generalized median graph. Graph data structure was defined by the spatial proximity of pairs of neighboring CTs and the best fit arrangement was determined for a given sample set [43]. The best fit arrangement represents the spatial arrangement of CTs and the association between a pair of CTs for a given sample set. The best fit of each sample

set can be used to compare the difference between samples from various cell line and/or states.

However, these approaches have few limitations. For example, if we only investigate the radial distribution, CTs that have the same distribution may actually locate in different octants in 3D space. Also, the CT proximity relationships and probability maps based on the relative distance of CTs only allow comparison of the association between two or more chromosome homologs, but do not allow analysis of the spatial organization of an individual chromosome homolog. Table 2.1 briefly summarizes existing algorithms that have been developed for analysis of CTs. Our proposed approach is able to assess the specific (x, y, z) coordinate of each CT within the nuclear space while also allowing precise quantification of their radial distribution, relative distance, and shape information.

Table 2.1. Summary of different approaches for analysis of chromosome territories.

	Approach	**Specific (x,y,z) coordinate**	**Radial distribution**	**Relative distance**	**Shape information**
Cremer et al.	Radial distribution	×	✓	✓	×
Berger et al.	Respect to nucleolus	×	×	✓	×
Brianna et al.	Neighborhood composition	×	×	✓	×
Zeitz et al.	Probabilistic model	×	✓	✓	×
This study	SPHARM	✓	✓	✓	✓

2.7 Motivation and Innovation

Aneuploid chromosomes (gain or loss of chromosomes) are a frequently observed trait in cancers. Researchers suggest that the alterations in gene expression mostly ensue in aneuploidy. For example, Kemeny indicates that the CT volume as well as radial positioning of some CTs are affected by the extra copy of chromosome 21 in chorionic

villus cell nuclei [44]. However, very limited and sometimes conflicting information is accessible to the connection between CT positioning and aneuploidy [45]. Cremer at al. evaluated eight tumor cell lines and found alteration in CT positioning in seven of them [8]. Nonetheless, others have suggested that the 3D-position of CTs remains similar to that of their endogenous normal homologues for artificially introduced aneuploid chromosomes [46]. Radial nucleus arrangement is a popular approach to study CT organization. However radial distribution characterization of CTs is unable to distinguish between CTs that have the same radial distance but lie within different octants in 3D space.

Here we present a modeling framework that allows determination of the 3D position of chromosome territories in a coordinate system (i.e., (x, y, z) value) with respect to a single chromosome within the nucleus as a predefined landmark. Our approach not only provides the specific coordinate position of a CT within an octant of 3D space, but also permits the investigation of the popular radial localization. Also, the unique coordinate position in 3D provides unambiguous determination of the local neighborhood of each chromosome that allows comprehensive study in proximity patterns of CT arrangement. The surface of nuclei and enclosed CTs from FISH signals are reconstructed using spherical harmonics which can precisely describe the surface shape of nuclei and CTs. As a result, our approach is able to provide specific position and shape information of CTs and nuclei. Moreover, based on the shape information acquired from spherical harmonics, we are able to perform cell-based, orientation and sized normalization; therefore, enabling comparison across cells and/or cell populations.

2.8 Spherical Harmonics

With the development in imaging techniques such as microscopy, laser scanning, and computer tomography, the acquisition of three-dimensional morphological data for various structures is now easier and faster. This kind of data brings up the needs of analytical techniques that are capable of quantifying complex 3D shapes. Spherical harmonics is well-suited for 3D modeling and has been applied to several fields including computer vision, computer graphics, and also medical image analysis. For example, a spherical harmonic model was applied to cardiac MRI analysis [47], and shape analysis of brain ventricles [48]. Spherical harmonics are series of orthogonal functions defined on the surface of a sphere and considered as the extension of Fourier techniques from two dimensions to three dimensions. A Fourier series can model a simple closed 2-D contour with Elliptic Fourier Descriptor [49]. Figure 2.6 gives an example of a simple contour (A). The shape of the square can be described with a radius function $r(\theta)$ that takes the angle in the polar coordinate system and gives the distance from origin to the boundary of the contour. The radial function can be formulated using Fourier series:

$$r(\theta) = \frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n \cos n\theta + b_n \sin n\theta), \quad (2.1)$$

where the Fourier coefficients a_n and b_n are defined by Fourier transform

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} r(\theta) \cos n\theta \, d\theta \text{ and} \quad (2.2)$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} r(\theta) \sin n\theta \, d\theta. \quad (2.3)$$

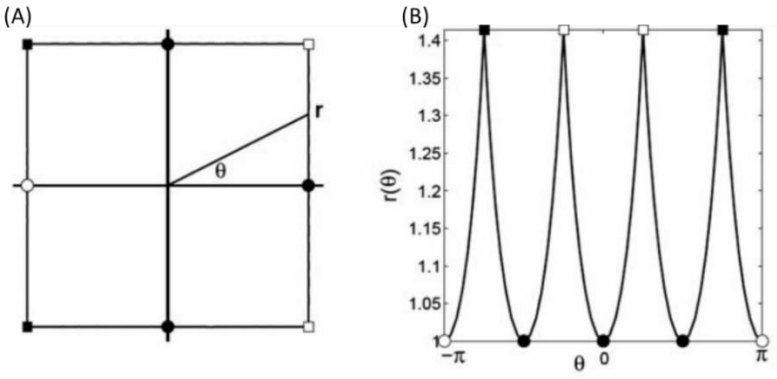

Figure 2.6. (A) A simple closed contour, square, and (B) a radius function $r(\theta)$ that describes the distance from the center of (A) to the boundary at different angular parameter, θ (figure from [50]).

In some cases, the 2D closed contour may not be able to be described by radius function for a given origin. Figure 2.7 gives three examples for such scenarios, the "L" shape contour has no value at angle θ for radial function and more than one radius value for $r(\theta)$ in the "C" and "T" shape contours. Thus, instead of using the radial function, two parametric functions $x(\theta)$ and $y(\theta)$ are used to represent the x- and y- Cartesian coordinates of the contour. The parameter θ now is defined as the arc length along the contour relative to an arbitrary origin on the contour. These two functions also can be expanded in form of Fourier series as

$$x(\theta) = \frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n \cos n\theta + b_n \sin n\theta) \text{ and} \quad (2.4)$$

$$y(\theta) = \frac{c_0}{2} + \sum_{n=1}^{\infty}(c_n \cos n\theta + d_n \sin n\theta), \quad (2.5)$$

where the Fourier coefficients, $a_n, b_n, c_n, and\ d_n$ can be calculated using the Fourier transform in equations (2.2) and(2.3).

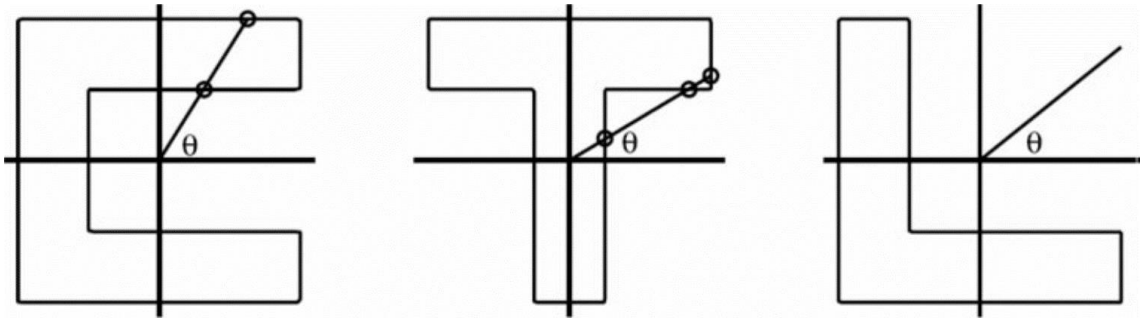

Figure 2.7. Examples of 2D closed contour that cannot be represented by radius function for given origins (figure from [50]).

As shown in Figure 2.8, extending the one radius function into two Cartesian coordinate functions with one parameter, and arc length, improve the shape descriptor such that it is able to represent an arbitrary shaped contour. Note that the parameter arc length is normalized to 2π, which is the circumference of a unit circle.

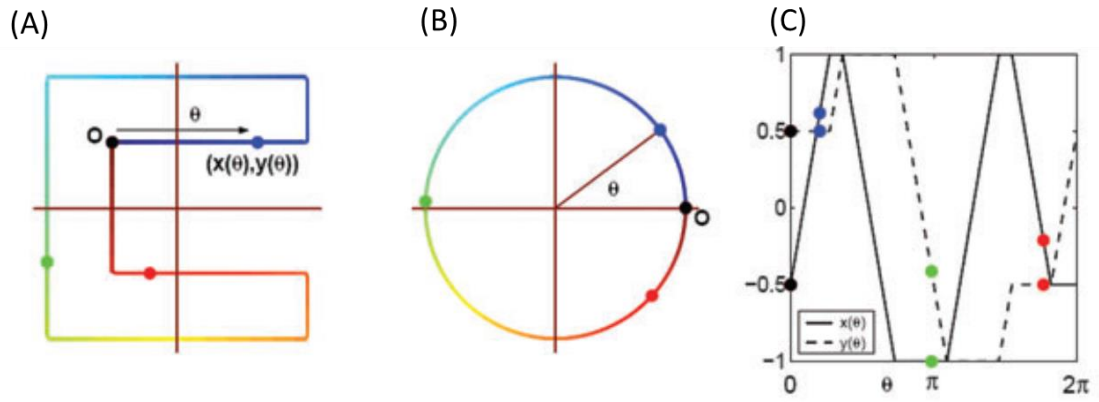

Figure 2.8. Example of (A) an arbitrary shaped closed contour that is parameterized on to (B) a circle that measure the underlying arc length. The contour is described by two functions $x(\theta)$ and $y(\theta)$ based on the arc length, θ (figure from [50]).

Finally, the 2D Fourier shape representation can be extended to 3D for modeling enclosed 3D surfaces. A radial function in 3D, $r(\theta, \emptyset)$, is based on parameters θ and $\emptyset$ representing the polar ($\theta \in [0, \pi]$) and azimuthal ($\emptyset \in [0,2\pi)$) coordinate respectively. Similar to 2D Fourier formulation, the radial function in 3D indicates the distance from a specified origin to each point on the surface and can be decomposed in terms of a Fourier

series and transform. However, like the 2D case, this radial parametrization has its limitation which cannot define the points on the surface that have a zero or multiple radial values. Brechbuhler *et al*. introduced an extended method called SPHARM that describes arbitrarily shaped simply connected objects with three explicit functions, $x(\theta, \emptyset)$, $y(\theta, \emptyset)$, and $z(\theta, \emptyset)$. Different approaches have been published for mapping the surface of the objects onto a unit sphere [50], [51], but with the same goal of minimizing area and topology distortion.

The three explicit functions that represent the Cartesian coordinates $x(\theta, \emptyset)$, $y(\theta, \emptyset)$, and $z(\theta, \emptyset)$ can be formulated as Fourier spherical harmonic (SPHARM) basis functions [51]

$$Y_l^m(\theta, \phi) \equiv \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)}} P_l^m(cos\theta) e^{im\phi} \ , \tag{2.6}$$

where $P_l^m(cos\theta)$ are the associated Legendre polynomials defined by the differential equation

$$P_l^m(x) = \frac{(-1)^m}{2^l l!}(1-x^2)^{m/2} \frac{d^{l+m}}{dx^{l+m}}(x^2-1)^l. \tag{2.7}$$

There are an unlimited number of spherical harmonics basis functions, $Y_l^m(\theta, \phi)$, that can be calculated at degree, l, and order, m, but practically, the basis functions are computed up to a user defined band L_{max}. l and m are integers with $0 \leq l < L_{max}$ and $|m| \leq l$. Figure 2.9 shows the shape of the spherical harmonics at the first five degrees.

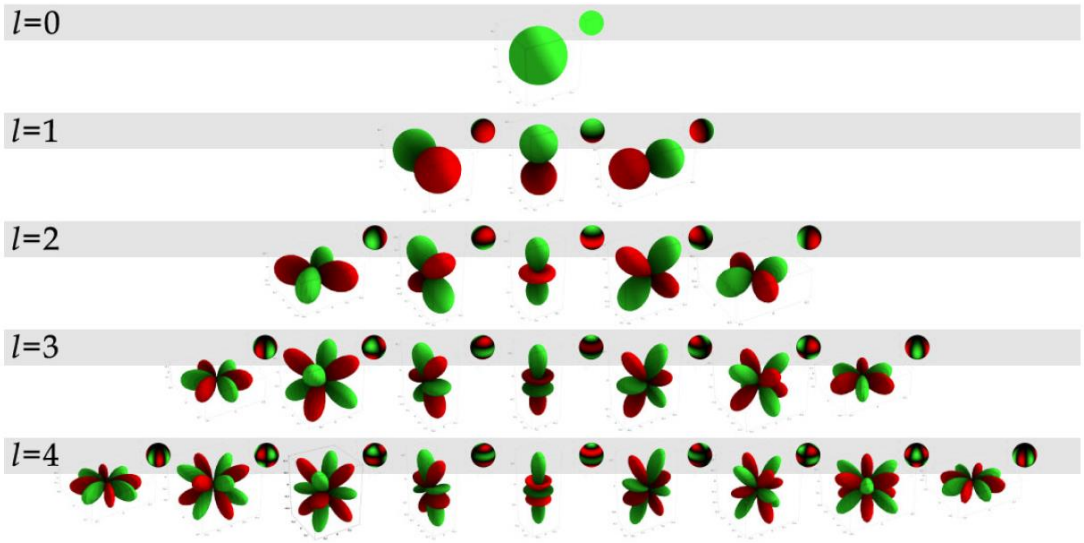

Figure 2.9. Demonstration of the first five spherical harmonics as unsigned spherical functions by distance from the origin and color on a unit sphere. Green indicates positive values and red represents negative values (figure from [52]).

Each of the three functions ($x(\theta,\emptyset)$, $y(\theta,\emptyset)$, and $z(\theta,\emptyset)$) is decomposed explicitly into a set of spherical harmonics basis functions as

$$x(\theta,\phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} C_{lx}^{m} Y_l^m(\theta,\phi), \tag{2.8}$$

$$y(\theta,\phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} C_{ly}^{m} Y_l^m(\theta,\phi), \text{ and} \tag{2.9}$$

$$z(\theta,\phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} C_{lz}^{m} Y_l^m(\theta,\phi). \tag{2.10}$$

C_{lx}^{m}, C_{ly}^{m}, and C_{lz}^{m} are the coefficients corresponding to spherical harmonic, Y_l^m. Each of XYZ coordinates is described in terms of individual spherical harmonic function with independent coefficients. Equations (2.8), (2.9), and (2.10) can be bundled into a vectorized function as

$$V(\theta,\phi) = (x(\theta,\phi), y(\theta,\phi), z(\theta,\phi))^T = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} C_l^m Y_l^m(\theta,\phi)\,, \tag{2.11}$$

where $C_l^m = (C_{lx}^m, C_{ly}^m, C_{lz}^m)^T$ is the coefficient matrix and $V(\theta,\phi) = (x(\theta,\phi), y(\theta,\phi), z(\theta,\phi))^T$. The coefficients could be estimated using the least-square algorithm and the surface is then reconstructed with these coefficients.

Chapter 3 – Method and Materials

3.1 Experimental Sample

The hES cell line WA09 (H9) at passage 29 was obtained from WiCell Research Institute (Madison, WI), and the undifferentiated cells were cultured on a mouse embryonic fibroblast feeder layer, strictly following the culture conditions suggested by WiCell, i.e., sub-culturing once a week; over 14 months until passage 87. In order to assess acquired aneuploidy over serial passaging, cultures from every other passage were monitored for undifferentiated cells using a rapid staining assay for alkaline phosphatase activity and/or FACS analysis of dual label SSEA1/SSEA4, and Oct4 as recommended by WiCell. In addition, differentiated cells were manually removed from culture during passaging using the protocol suggested by WiCell. Both morphologically and physiologically, the differentiated cells are distinct from the pluripotent hES cells. In this study, we used only the undifferentiated hES cells, since changes in nuclear architecture have been noted in differentiated cells [53]. These changes are indeed related to changes in the distinctive localizations for chromosome regions and gene loci with a role in pluripotency. Wiblin *et al*. [53] found that 12p, a region of the human genome that contains clustered pluripotency genes including NANOG, has a more central nuclear localization in embryonic stem cells than in differentiated cells. Also a smaller proportion of centromeres located close to the nuclear periphery in hES cells compared to differentiated cells. Therefore, differentiated cells were removed from the culture. For each passage cell karyotype was analyzed by conventional G-banding, FISH (using centromeric probes for chromosome 12 and 17) and SKY methods. The diploid WA09 cells progressively became aneuploid between passages 35 to 87 by gaining an extra copy of chromosome 12. With increase in passage the

percentage of trisomy-12 cells increased and reached approximately 70% by passage 68 and then remained constant until the culture was terminated at passage 87. 27 diploid and 29 trisomy-12 nuclei were imaged from the passage 73 cell population.

3.2 FISH Slide Preparation and Image Acquisition

Chromosomal abnormalities such as an increase or decrease in the number of chromosomes are visualized using chromosome specific DNA probes via multi-color fluorescence in situ hybridization (FISH). In FISH, a DNA probe, that contains complementary sequence to the target chromosome, is hybridized to the chromosome within the interphase nuclei. The location of the target chromosome can then be visualized directly or indirectly through fluorescence immunocytochemical techniques.

3.2.1 3D-FISH with Whole Chromosome Painting

In order to study the three-dimensional relationship of chromosome territories, we use 3D-FISH with whole chromosome painting probes to determine the spatial positioning of the entire CT. In Table 3.1, all the chemicals used in the cell fixation and 3D-FISH are listed with company and/or distributors information. Fixation and 3D-FISH pretreatment of sample slides was performed using hES cells interphase nuclei at passage 73 by following the protocols from Cremer et. al., [7].

Table 3.1. Chemicals

Chemicals	Part number, company, distributer
SSC (buffer solution mixture of sodium chloride and sodium citrate)	30-805850, Abbott Molecular Inc., IL, USA
Phosphate Buffered Saline (PBS)	46-013-CM, Mediatech Inc., VA, USA
Tween	H5152, Promega, WI, USA
Magnesium Chloride ($MgCl_2$**)**	5958-04, Avantor Performance Materials, PA, USA
Hydrochloric Acid (HCl)	H613-45, Avantor Performance Materials, PA, USA
Formamide	F7503-1L, Sigma-Aldrich, MO, USA
Triton-X	X100-500ML, Sigma-Aldrich, MO, USA
Pepsin	P7000-25G, Sigma-Aldrich, MO, USA
Glycerol	49767-250ML, Sigma-Aldrich, MO, USA
Paraformaldehyde (PFA)	15710, Electron Microscopy Sciences, PA, USA
Vectashield with DAPI	H-1200, Vector Laboratories, CA, USA

Protocol for Cell Fixation

1. Briefly rinse the slide with cells grown to sub-confluency in three changes of 1xPBS at 37°C.
2. Fix in 4% PFA in 1xPBS (pH 7.0) at room temperature (RT) for 10 min. During the last minute, add a drop of 0.5% Triton X-100/PBS.
3. Wash in 1xPBS with 0.01% Triton X-100/PBS at RT for 3x 3min.
4. Incubate in 0.5%Triton X-100/PBS at RT for 15 min.
5. Incubate in 20% glycerol in 1xPBS at RT for 60 min.
6. Freeze by dipping the slide into liquid nitrogen for 30 sec and thaw on a piece of paper towel. As soon as the frozen layer disappears, put the coverslip back into 20% glycerol/PBS and repeat four times.
7. Wash in 1xPBS for 3x 10 min.
8. Incubate in 0.1 N HCl for 5 min at RT.
9. Incubate in 2x SSC for 2x 3 min.

10. Incubate in 50% formamide (pH = 7.0)/2x SSC for 1 hour at RT before proceeding with pepsin digestion.

Fixed samples prior to pepsin treatment maybe stored in 50% formamide/2x SSC at +4°C. Fixed samples may be stored for a period of at least 3-4 months prior to hybridization. Fixedsamples were then treated with pepsin for efficient hybridization.

Protocol for Treatment with Pepsin

1. Equilibrate slides (kept in 50% formamide/2× SSC) in 2× SSC at RT for 2 min.
2. Equilibrate slides in 1× PBS at RT for 3 min.
3. Incubate in pepsin (0.005% in 0.01 N HCl): warm 0.01 N HCl to 37°C in a bottle and add pepsin (30–50 µL of 10% stock solution (ddH_2O)) just before use, shake well, and pour into a Coplin jar (7-200-011, American Educational Products, CO, USA). Incubate in pepsin for 3–5 min. (5 min)
4. Incubate in 1× PBS/50 mM $MgCl_2$ to inactivate pepsin, RT for 2× 5 min.
5. Postfix in 1% paraformaldehyde/1× PBS, RT for 10 min.
6. Wash in 1× PBS, RT for 5 min.
7. Wash in 2× SSC for 2× 5 min, then return slides to 50% formamide/2× SSC for at least 1 h before hybridization.

3.2.2 Probe Denaturation and Setup of Hybridization

Cells were labeled using DAPI (4’,6-diamidino-2-phenyl indole dihydrochloride) as the nuclear counterstain, and whole chromosome painting probes (Applied Spectral Imaging, USA) for chromosomes X, 8 and 12 tagged with FITC (fluorescein isothiocyanate), Aqua, and TRITC (tetramethylrhodamine B isothiocynate), respectively.

A simultaneous denaturation of nuclear and probe DNA is recommended since it is optimal for retention of 3D morphology.

Protocol for Probe Denaturation and Setup of Hybridization

1. Place the hybridization mixture with dissolved probe on a coverslip (e.g., 6– 8µL per 18×18-mm coverslip).
2. Take a slide with cells out of the 50% formamide/2× SSC and quickly drain the excess of fluid off the slide.
3. Cover the target area of the slide on the coverslip with probe.
4. Wipe off the excess fluid around the coverslip and seal with hybridization cover (712222, Grace bio-lab, OR, USA) . Protect from light during drying in case of fluorochrome-labeled probes).
5. Place slides on a hot block at 75°C to denature cellular and probe DNA for 7 min. Keep this time and temperature strictly.
6. Perform hybridization in a microsample incubator (157-30, SciGene, CA, USA) insert with wet paper towels to maintain humidity at 37°C for 3 days.

3.2.3 Washing and Detection

The quality of an antibody may vary depending on the batch; thus, the protocol for hybridization and detection was iteratively modified and optimized.

Protocol for Washing and Detection

1. Remove the slides from the humidified chamber and carefully remove the hybridization cover.
2. Transfer slides to a coplin jar containing 0.4xSSC. Wash slides in 0.4xSSC at 74°C ± 2°C for 5 min (pre-heat the coplin jar). Dip slides in washing solution II (4xSCC/0.1% Tween 20) for 2 min.
3. Mount hybridized area in antifade (Vectashield with DAPI) (around 10 micro liter).

3.2.4 Confocal Imaging Acquisition

Confocal images were acquired on an Olympus FluoView 1000™ system (Olympus Corp. of the Americas, Center Valley, PA, USA). Plan-Apochromat 60x oil immersion objective with 1.4 Numerical Aperture was used for imaging. Serial optical sections were acquired at a z-interval of 0.3-μm and pixel size of 0.059 μm/pixel. Digital zoom value of 7 was used while acquiring images of single nuclei. The optical images were acquired in the four-channel mode with laser wavelengths of 488, 458, 543, and 405 nm for chromosome X, 8, 12 and nuclei respectively.

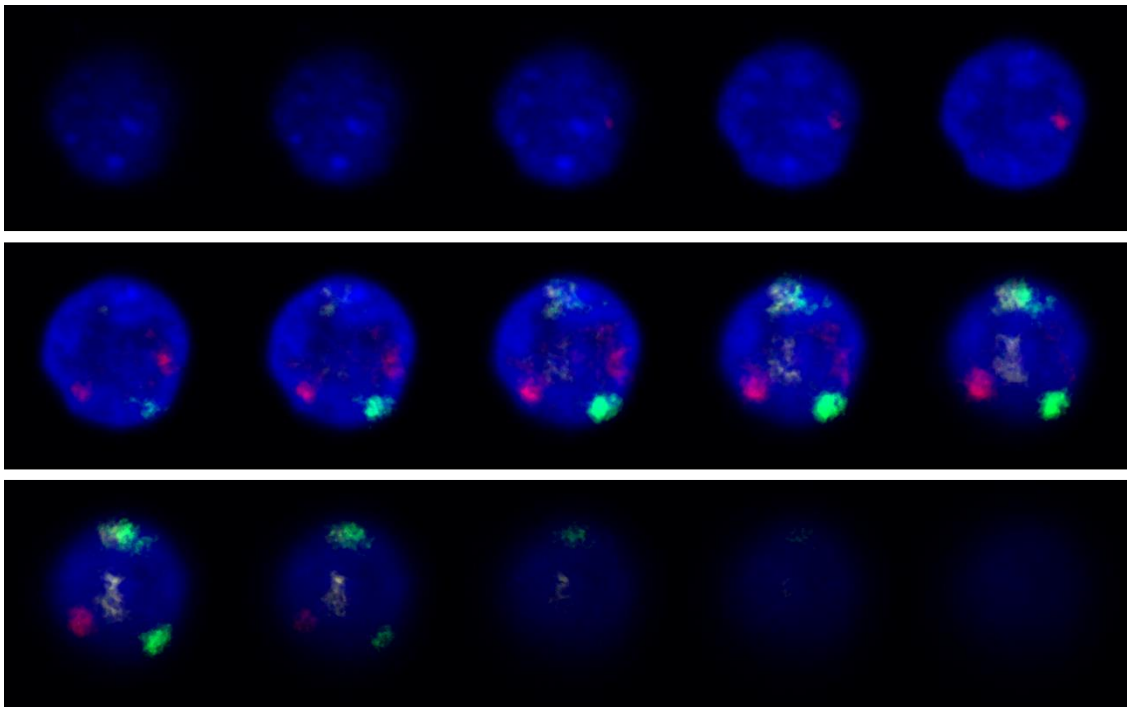

Figure 3.1. Montage image of one image stack of a nucleus

Each 3D image stack of a nucleus consists of multiple images at different depth along vertical axis (z-axis). The first image of the stack captures the top-most slice of the nucleus and the last image gives the bottom most slice of the nucleus. Figure 3.1 shows a montage of the optical slices for a representative nucleus image using 4-color confocal fluorescence microscopy. The image on top-left corner is the first image of the stack and the depth increases in order from left to right then top to bottom. The blue color channel shows the entire nucleus illuminated by fluorescence of DAPI. The green color channel captures the signal from the fluorescence of FITC (fluorescein isothiocyanate) tagged chromosome X. The cyan color channel detects the fluorescence light signal Aqua that represents chromosome 8. The TRITC (tetramethylrhodamine B isothiocynate) is attached to chromosome 12 and its fluorescence light is seen in the red color channel.

The merged image stack is been analyzed in four separated channels as shown in Figure 3.2. The merged image is split into four channels with color green, red, yellow, and blue representing chromosome X, 12, 8, and nucleus.

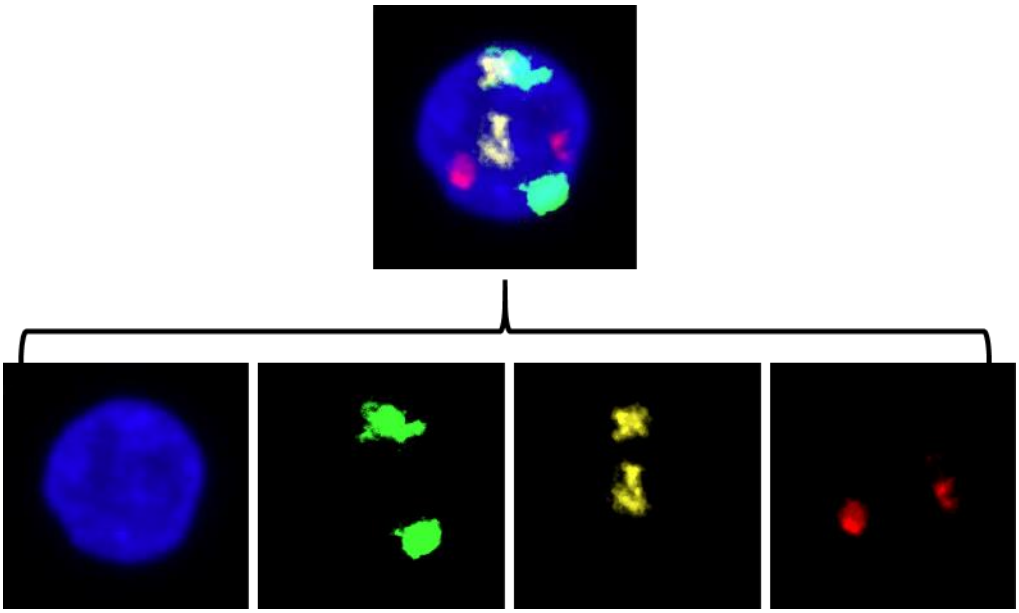

Figure 3.2 Split merged image into four color channels.

3.3 Image Segmentation

Images in each of the four channels are pre-processed to eliminate noise and segmented, to detect each homologous chromosome for further analysis.

3.3.1 Pre-processing

Every slice of an image stack for each channel is saved as an image array. A median filter, a nonlinear filtering technique that is widely use in image processing, finds the median over a sliding window with fixed size and assigns the median value to the center of the sliding window. Linear filtering techniques such as a mean filter that computes the average over a fixed size window could result in blurring and consequently lost edge information. On the other hand, median filtering performs better preserving edge

information in an image [54]. Each image in the array has size of 230x232 and is filtered with a median filter to eliminate undesired noise in the image array before being segmented. In this research, an empirically determined 20 by 20 pixels window is used for median filter to eliminate unwanted noise signal.

3.3.2 Binarization

The optical sections are scanned in a four channel mode at 230X232 pixels per frame using 12-bit pixel depth for each channel at a voxel size of 0.059 x 0.059 x 0.3 μm. Therefore, each image slice from each channel are gray scale images in size of 230X232 with intensity value range from 0 to 4095.

Each grayscale image is divided into two classes called foreground and background (transform values from 0 to 4095 to 0 and 1) by thresholding. Thresholding is one of the simplest image segmentation method. It examines each pixel value in an image and compares it to a threshold value. If the value of the current pixel is greater than the threshold value, it is assigned to foreground class with a value 1, whereas a pixel is considered as background with a 0 value if its value is less than the threshold value.

The intensity of fluorescent signal may vary across different samples and channels. Thus, a threshold value is calculated specifically for each image stack in each of the 4 channels. To determine the threshold value, first the slice that has the highest integrated intensity value is selected, and then the threshold value is computed using an iterative algorithm. Each pixel in the selected image slice is compared to a predefined initial value and classified to two groups, foreground and background. The initial value is determined by 40% of the maximum pixel value of the selected image. A new threshold value is

calculated by averaging the mean value of foreground and background. The new threshold value is now used for the next iteration of thresholding until the threshold value stops changing or the change is less than a tolerance value of 0.001. The output threshold value from the selected slice is then applied to all slices in the image stack for thresholding.

3.3.3 Region Detection

After the thresholding, only foreground pixels in the binary image are used for further analysis. Except for the nuclei channel which contains only one foreground object, the other channels contain two or more 3D chromosome objects in an image stack. Region labeling is used to segment the 3D objects (i.e., nucleus and individual chromosomes) in each image stack. The 3D foreground voxels (volume pixel) in an image stack that are connected together are assigned to a unique label. The voxels that have the same label are considered in one distinct region. This algorithm runs through all the foreground voxels and checks connectivity with its neighboring voxels in the same slice as well as the voxels in the image slice directly above and below the current voxel.

Two or more voxels can be connected via a common edge, a common vertex, or a common face. This means a voxel can potentially be connected to 6, 18, or 26 neighboring voxels respectively. In our case, we consider the distinct region is connected by a vertex, i.e., a 26-connected neighborhood is used for region detection. For each foreground voxel, all the corresponding 26 neighbor voxels are examined to see if they are connected to the current voxel. All the foreground voxels that share a common vertex to the current voxel are considered as the same group and assigned to the same label. Figure 3.3 shows the neighboring voxels of current voxel in 3D. The blue voxels indicate the voxel that have

common face, common edge, and common vertex to current voxel (red) with 6, 18, and 26 connecting voxels respectively.

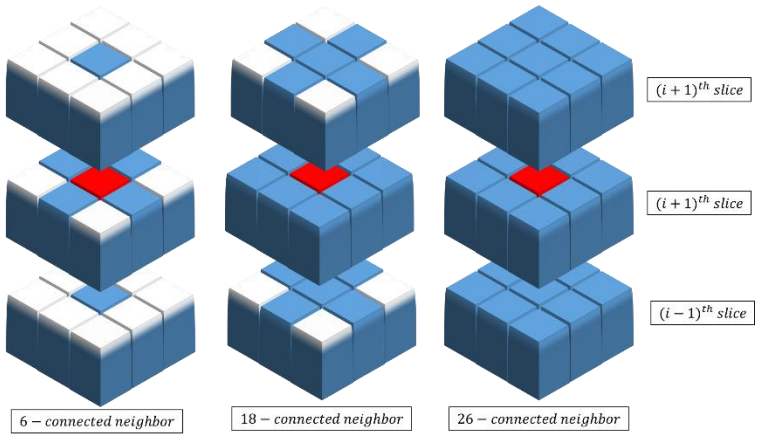

Figure 3.3. 6, 18, and 16 -connected region detection neighbor.

The region labeling step is accomplished with Matlab function 'bwlabeln' that outputs the 3D binary volumetric images of nuclei and individual chromosomes. For a normal nucleus, there is one 3D binary volumetric image for nucleus and two each for chromosome X, 8, and 12. Aneuploid nuclei have one extra 3D binary volumetric image for chromosome 12.

3.4 Estimation of 3D Surface using Spherical Harmonic Modeling

3D images (z-stacks) from confocal fluorescence microscopy were pre-processed, and 3D nuclear and CT objects were segmented using gray level thresholding and 3D region labeling algorithms. Surface shapes for nuclei and CTs were estimated using spherical harmonics.

The volumetric structure of nuclei and the 3D morphology of chromatin were estimated by spherical harmonic modeling (SPHARM) described by Brechbühler [51]. SPHARM

modeling consists of three steps: (1) parametrization, mapping the vertices from Cartesian coordinate system (x, y, and z coordinate) onto a unit sphere, (2) expansion that calculates the SHPARM descriptor (or set of coefficients), and (3) reconstructing the 3D object for a given set of vertices. The SPHARM descriptor is computed using the SPHARM-MAT toolbox developed by Shen[55] as shown in Figure 3.4.

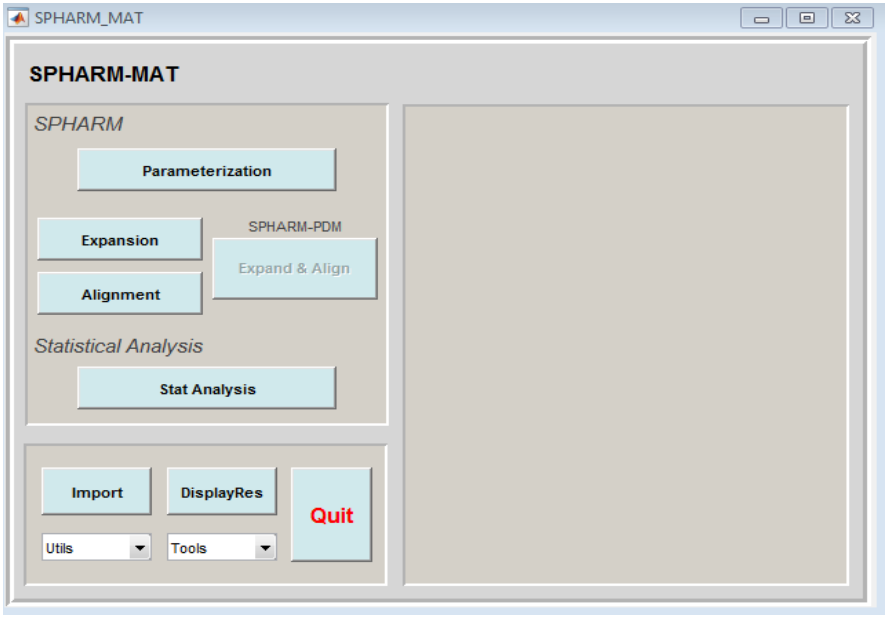

Figure 3.4. The graphical user interface of SPHARM-MAT toolkit [55].

The SPHARM-MAT toolbox provides methods to perform parameterization, registration, and statistical analysis. It also includes some other utilities, such as topology fix for the input data, computing average shape, and displaying the object.

The input data format for SPHARM-MAT is in a pre-defined "*.bim" format that contains binary volumetric data and voxel size. After the image segmentation, our FISH signals of CT and nucleus are transformed to volumetric objects. The voxel size is set to 1 x 1 x 5, corresponding to resolution of the FISH signal (0.059 x 0.059 x 0.3 μm).

3.4.1 Spherical Parameterization

First step of SPHARM modeling is to find the continuous and uniform mapping from a 3D object onto a unit sphere. There are three types of distortions when mapping the transformation: area distortion, length distortion, and angle distortion. The ideal mapping is isometric (length preserving), and it happens if and only if both the angle and area are preserved. However, isometric mapping occurs only in certain special scenarios. For instance, when mapping a cylinder onto a plane. Therefore, equiareal (area preserving) mapping is an equitable choice for sphere mapping. Since an uncontrolled equiareal mapping usually causes unwanted angle and length distortions, minimization on length and angle distortions is required. The mapping results in a bijective mapping from each point $v(x, y, z)$ on the surface to spherical coordinates θ and ϕ as shown in Figure 3.5.

The toolkit implements two methods to complete the task. The first one is originally proposed in SPHARM paper by Brechbühler [51]. It creates a one-to-one mapping from an object to a unit sphere with an equal area via minimizing the area distortions as an optimization problem and this method is implemented in an earlier version of the toolkit called SPHARM-PDM [56]. The second method is proposed by Shen[57], and extends the traditional method with control of both area and length distortions (CALD). Unlike the traditional method, which is only applicable to volumetric images, CALD is applicable to general triangular mesh data sets. Also, it uses a two stage method to ensure that both area and length distortions are minimized.

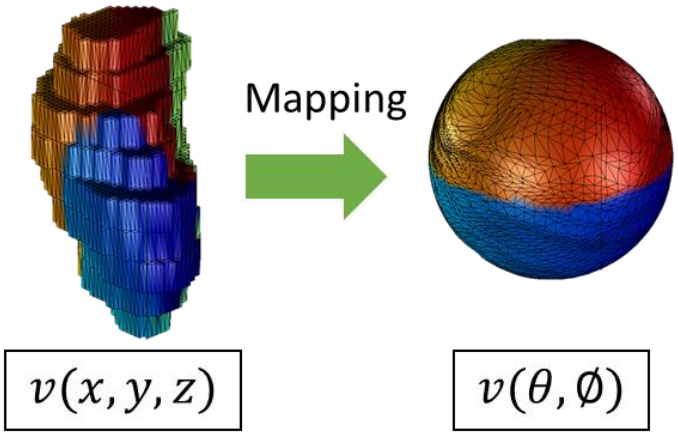

Figure 3.5. Continuous mapping from Cartesian (x, y, z) coordinates onto spherical coordinates (θ, ϕ).

CALD comprises of two steps: local mesh smoothing and global smoothing. It starts from an initial parameterization and then performs local mesh and global smoothing alternatively until the desired result is achieved. Local mesh smoothing moves the vertices on the unit sphere to minimize the area distortions at a local sub-mesh while restricting any line distortions. This smoothing step only alters the vertices distribution on the sphere but does not change the object mesh and mesh topology. The global smoothing then equalizes the area distortion distribution of the entire sphere.

CALD method improves the initial parametrization from Brechbuhler's [58] method by applying local mesh smoothing and global smoothing iteratively. In order to use the SPHARM toolkit, the image stacks of nuclei and individual chromosomes need to be converted to a format that is compatible with the toolkit. CALD method takes mesh input and the SHPARM-MAT toolkit automatically converts the volumetric data to mesh surface by connecting the vertices of the voxels and dividing the surfaces into triangles. Therefore, the 3D binary volumetric images of nuclei and chromosomes as described above are ready

for applying the spherical parametrization. Details of the CALD mapping technique are described by Shen *et al*.[57].

3.4.2 SPHARM Expansion

In the previous sections, the 3D images (z-stacks) from confocal fluorescence microscopy have been pre-processed, segmented using gray level thresholding, labeled with 3D region detection algorithms, and the 3D binary volumetric images of nuclei and chromosomes are mapped onto a unit sphere.

Next the surface was expanded into a set of spherical harmonic basis functions with coefficients for each spherical harmonic. The shape of the modeled objects is described by the sum of spherical harmonics in the different dimensions (degree) weighted by the coefficients:

$$V(\theta,\phi) = (x(\theta,\phi), y(\theta,\phi), z(\theta,\phi))^T = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} C_l^m Y_l^m(\theta,\phi), \quad (3.1)$$

where $C_l^m = (C_{lx}^m, C_{ly}^m, C_{lz}^m)^T$ is the coefficient matrix that contains the corresponding spherical harmonic with degree l and order m, Y_l^m. Each of XYZ coordinates is described in terms of individual spherical harmonic function with independent coefficients $C_{lx}^m, C_{ly}^m, \text{ and } C_{lz}^m$. Therefore, for a 3D object, the vertices on a surface $V(\theta,\phi)$ can be independently decomposed into equation (3.2):

$$x(\theta,\phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} c_{lx}^m Y_l^m(\theta,\phi), \quad (3.2\text{-a})$$

$$y(\theta,\phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} c_{ly}^m Y_l^m(\theta,\phi), \text{ and} \quad (3.2\text{-b})$$

$$z(\theta,\phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} c_{lz}^m Y_l^m(\theta,\phi). \quad (3.2\text{-c})$$

The Fourier coefficients C_l^m can be estimated by solving a linear system. For example, to solve the estimation of coefficient of x-axis, $\hat{c}_{lx}^m$, we take the equation (3.2-a) and the linear system can be formulated as follow:

$$\begin{bmatrix} y_{1,1} & \cdots & y_{1,k} \\ \vdots & \ddots & \vdots \\ y_{n,1} & \cdots & y_{n,k} \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \tag{3.3}$$

where x_i is the value function of the input object sample such that $x_i = x(\theta, \phi)$, for $1 \leq i \leq n$. $y_{i,j}$ is the spherical harmonic basis function, where $y_{i,j} = Y_l^m(\theta, \phi)$, $j = l^2 + l + m + 1$. C is the corresponding coefficient. The total number of coefficients, k, for one axis can be up to $k = (L+1)^2$, and is directly determined by the user determined maximum degree, L. The total number of linear equations, n, in the system is determined by number of data points.

The SPHARM-MAT toolbox uses the Least-square estimation to find the approximate solution of the linear system stated above in equation (3.3). The method of least square refers to finding an overall solution that minimizes the sum of the squares of the residuals resulting from a set of equations. The residual is the difference between observed value and the output value from the estimation. Typically, the least-square method is suitable for an overdetermined system, which has a higher number of equations than unknown variables. In our case, total number of equations is defined by the number of data points (or vertices) of the 3D object and it can range from 3000 up to 50000 points depending on the size of the object. The unknowns are the coefficients of spherical harmonic its total number is based on user-defined maximum degree, L. The number of coefficients is $(\mathrm{L}+1)^2$. For example, we use a maximum degree of 12 for this study and the total number of coefficients is 169. We can control the number of coefficients and restrict our system to

over-determine. Therefore, least-square estimation is used to find the estimated coefficients, $\hat{c}_l^m = (\hat{c}_{lx}^m, \hat{c}_{ly}^m, \hat{c}_{lz}^m)^T$. Since $\hat{c}_l^m$ is the estimation of C_l^m, where $\hat{c}_l^m \equiv C_l^m$, the reconstructed 3D surface, $\hat{V}(\theta, \phi)$, is established as

$$\hat{V}(\theta,\phi)\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\hat{c}_l^m Y_l^m(\theta,\phi) \approx V(\theta,\phi)\,. \tag{3.4}$$

3.4.1 Surface Reconstruction

After finding the estimated coefficients for the input 3D object, the 3D surface of the object can be reconstructed by inserting a pre-defined nearly uniform sampling of spherical surface, $u(\theta, \phi)$, into equation (3.2). The SPHARM coefficients derived from a 3D object contain shape information of the object. It estimates a closed surface of the input object; hence, it is capable of reconstructing the 3D object surface with any input spherical vertices.

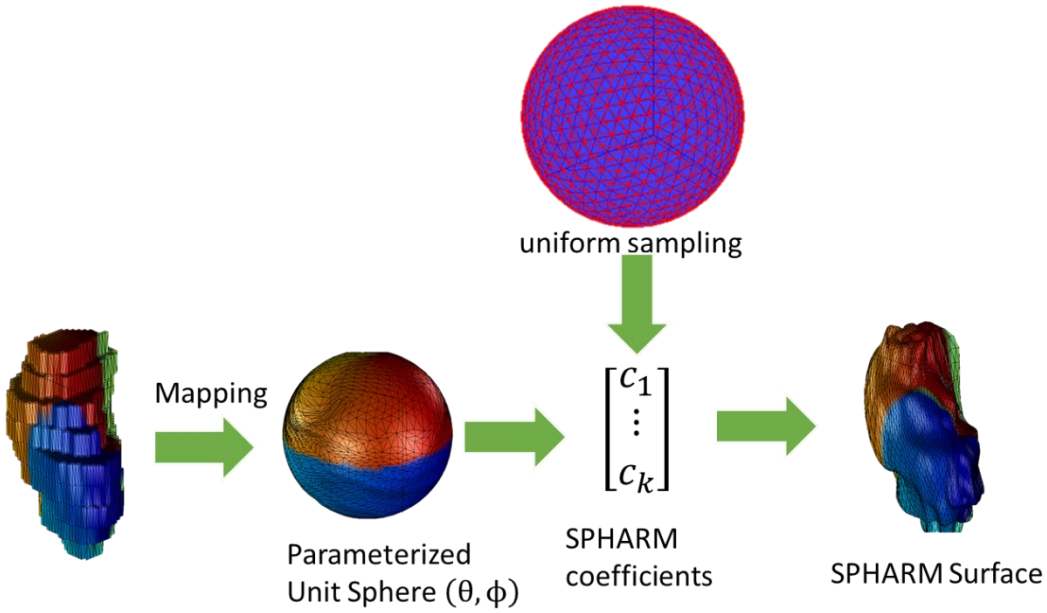

Figure 3.6. Schematic diagram of 3D surface modeling using SPHARM.

Figure 3.6 illustrates the SPHARM modeling approach with a schematic diagram. A 3D object is mapped onto a unit sphere then used for computing the SPHARM coefficients.

A set of uniform distributed sphere point set is coupled with the coefficients of the 3D object and a smooth 3D surface is reconstructed.

3.4.2 Selection of Degree for SPHARM Expansion

For a SPHARM estimation, there is one parameter that has to be specified by user, that is, the maximum degree. Choosing a proper degree can give an accurate estimation in a timely manner. Figure 3.7 gives two examples of 3D object reconstructed using SPHARM estimation with different degrees. The degree for SPHARM expansion directly affects the performance in the surface estimation of a 3D object. The higher the degree, the more detailed the reconstructed surface, $\hat{V}(\theta, \phi)$. However, using a higher degree also requires a larger number of coefficients that need to be estimated.

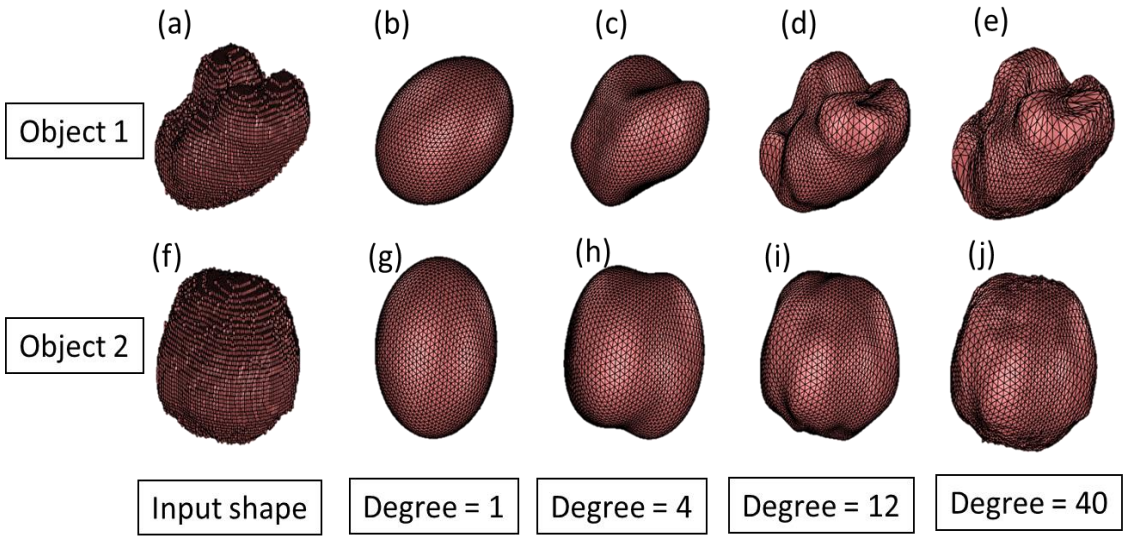

Figure 3.7. Reconstructed 3D surface using SPHARM estimation with different degrees. (b) - (e) show the reconstructed 3D surface of object 1 in (a) with degree equals to 1, 4, 12, and 40. (g) – (j) show the same result to (f).

There are two problems that have to be considered when selecting SPHARM degree. The first tradeoff for a higher degree estimation is an extensively higher computational

expense. The number of coefficients is determined by the pre-defined user selected degree. Also, the estimation, $\hat{V}(\theta, \phi)$, shown in equation (3.4) has to apply to each of the three x, y, and z axis. This means that the number of coefficients for a SHPARM expansion is tripled for a 3D object estimation. Thus, the number of coefficients equals to the number of unknowns that need to be estimated. For example, for a degree of 8 SPHARM estimation, the total number of coefficients is $(8 + 1)^2 * 3 = 237$ and in case of degree of 12, there are 507 unknowns that have to be solved. The other factor that has to be taken into consideration is overfitting. A higher degree has a larger number of coefficients to describe the 3D object surface. Nonetheless, an accurate estimation is not necessarily achieved with a higher degree for SPHARM. For example, Figure 3.8(a) shows a 3D surface of a nucleus sample and Figure 3.8 (b) is reconstructed surface of Figure 3.8 (a) with degree of 12. Figure 3.8 (c) shows an example of overfitting the surface of object in Figure 3.8 (a) at degree of 40.

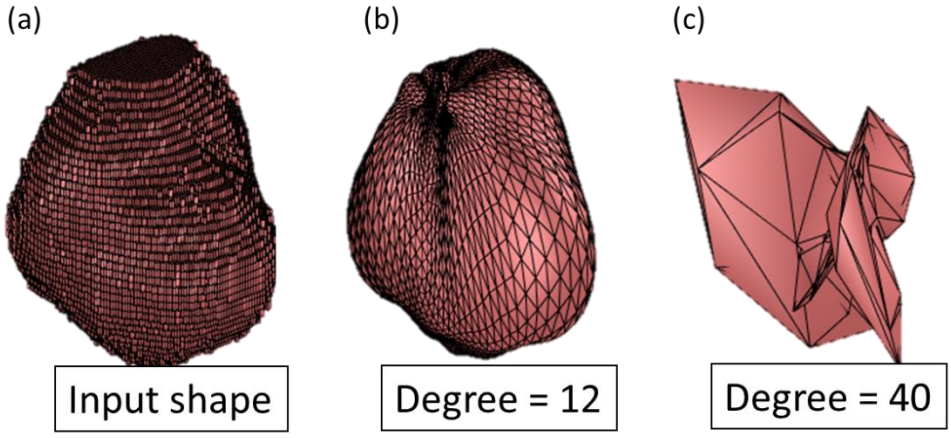

Figure 3.8. Demonstrate overfitting problem. (a) The input 3D object and reconstructed 3D surface using SPHARM estimation with degree of 12(c) and 40(c).

To determine a degree that is suitable for our data set, we evaluate the performance by calculating the root-mean-square error (RMSE) between input vertices, $V(\theta, \phi)$, and the estimated vertices, $\hat{V}(\theta, \phi)$:

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(\hat{V}(\theta_k, \phi_k) - V(\theta_k, \phi_k))^2}, \quad (3.5)$$

where n is the number of vertices in input vertices and estimated vertices.

The goal here is to find a degree that is low and has small RMSE while not having overfitting at the same time. In order to do so, we first use four test objects to evaluate the performance of SPHARM shape descriptor with degree of 1, 2, 3, 4, 8, 12, 16, 25, 30, 35, and 40. To ensure the least square estimation works properly, SPHARM-MAT tool kit restricts the degree based on the condition that the total number of unknown is less than the number of equations (the number of data points of an object). Specifically, the number of equations has to be at least twice the number of unknown. Data from nuclei and CTs both have data points more than 3362 for degree of 40. As shown in Figure 3.9, four simulated 3D objects: sphere, ellipsoid, rotated ellipsoid, and deformed ellipsoid are used for the evaluation. To be sure that the degree selection is suitable for our data set, we also include 12 randomly selected nuclei objects and 15 CT objects from our sample set for the evaluation.

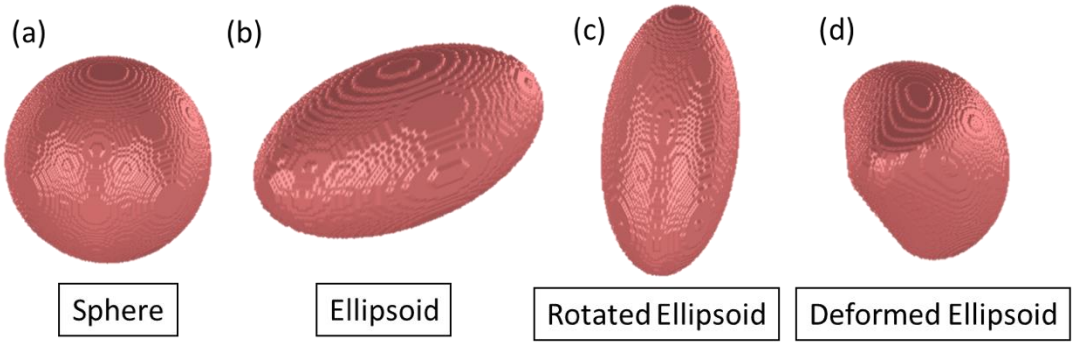

Figure 3.9. Test objects for evaluating the performance of SPHARM shape descriptor.

The result indicates that the RMSE of all test objects decrease with increasing of SPHARM degree and the actual RMSE value start to converge at degree of 12 as shown in Figure 3.10. Hence, the preliminary choice of degree for SPHARM descriptor is 12 to achieve a balance between fitting accuracy and computational complexity for this application under particular circumstances. The next step is to test to see if overfitting can happen when degree 12 is used to compute the SPHARM coefficients.

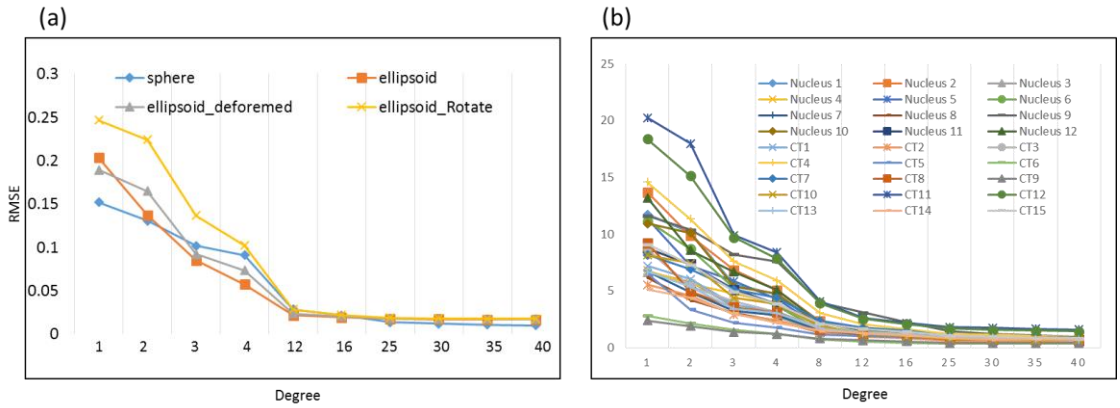

Figure 3.10 The RMSE of SPHARM reconstructed surfaces with different degrees using (a) simulated shapes and (b) 12 nuclei and 15 CTs from our samples.

The assessment for overfitting is done by computing the RMSE between resampled vertices of the object and the SPHARM estimated vertices of the object using the

coefficients derived from the original object. The process as shown in Figure 3.11 can be broken down to the following steps:

1. Parameterize closed object, $P_{vertices}(x, y, z)$, onto unit sphere as $P_{vertices}(\theta, \phi)$.
2. Solve for the SPHARM coefficients, c given $P_{vertices}(x(\theta, \phi), y(\theta, \phi), z(\theta, \phi))$.
3. Calculate face centroid of closed object, $P_{faces}\big(x(\theta, \phi), y(\theta, \phi), z(\theta, \phi)\big)$.
4. Use SPHARM coefficients from step 2 and the spherical coordinates from step 3 to estimate the face centroids, $\hat{P}_{faces}(x(\theta, \phi), y(\theta, \phi), z(\theta, \phi))$, using equation (3.2).
5. Compute the RMSE between $P_{faces}(x, y, z)$ and $\hat{P}_{faces}(x, y, z)$.

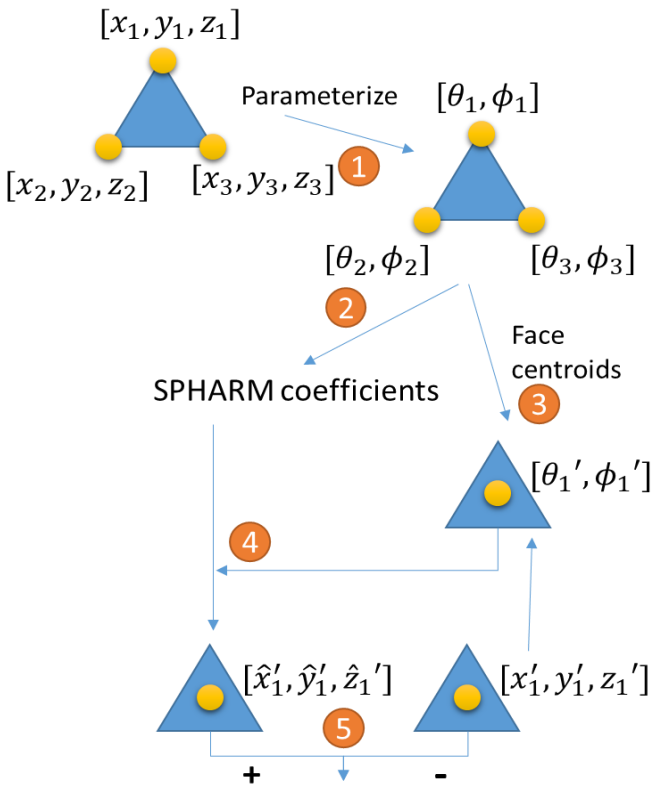

Figure 3.11. Process of evaluating RMSE with resample vertices.

We use the same sample objects from previous RMSE evaluation, compute the SPHARM coefficients, and conduct the RMSE assessment with the resampled vertices. From the result of RMSE assessment of resampled vertices, the RMSE of test objects stay stable at lower SPHARM degree and significantly increase at higher degrees. As shown in Figure 3.11, the RMSE of resampled vertices gives the same trend as the original objects at degree of 1, 2, 3, 4, 8, 12, 16, 25, and 30, but largely increases at a degree of 40 for a few nuclei objects and for few CT objects after degree of 30. Please note that for the object CT7 and CT3, the maximum degree tested was 30 since they didn't have enough data points to proceed to higher degree. At this point, the preliminary selection of SPHARM degree of 12 does not show tendency of overfitting and is computationally faster than higher degrees of (16-30). Thus, degree of 12 is selected for the 3D SPHARM modeling of our framework.

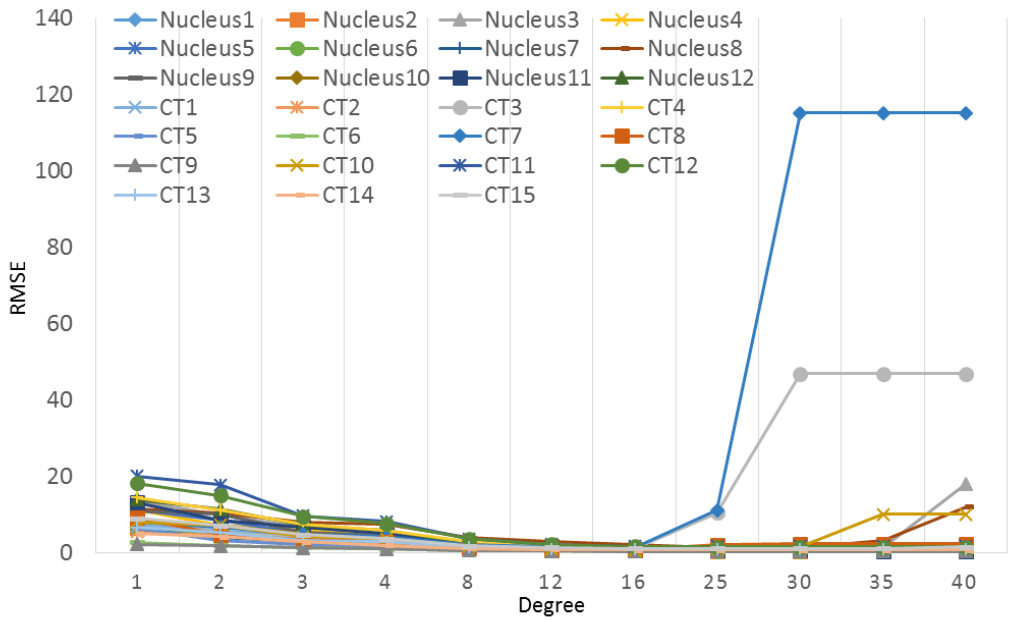

Figure 3.12. 12 nucleus objects and 15 CT objects RMSE with different degrees using resampled vertices.

3.4.3 Normalization

After reconstructing the 3D shape of nuclei and chromosomes, all the samples have to be normalized according to their size for further analysis. For each nucleus, we first move the center of the nucleus to the origin in Cartesian coordinate (0,0,0) and then normalize the nucleus and chromosome within it by the average size of the nucleus. The average size is computed by taking the mean of the distances of each vertex from the nucleus center.

3.5 Determination of Specific 3D Position in Space

Nuclei are typically found at different orientations based on how the cells fall on the slide during sample preparation. Thus, the imaged orientation of the nucleus varies across the different cell samples. This variance may affect the determined spatial localization of chromosomes inside the nucleus. This means when the nuclei have different orientation (e.g., 90^0, 180^0, etc.), the position of chromosomes may be presented differently even though their spatial organization is similar. Therefore, it is important to normalize the orientation to a common framework, and then align the samples to a fixed reference point when comparing the specific 3D position of CTs across different nuclei.

3.5.1 Common Reference Frame of Orientation

A common reference frame is established by aligning all the nuclei to a template based on the shape information provided by the SPHARM representation. The reasoning of aligning nucleus according their shape is that cells share similar shapes, and thus the SPHARM coefficients can be used to align the nuclei based on their shape. Rotation estimation between two objects is often done by comparing a set of corresponding points

extracted from both objects and the accuracy of the orientation estimation heavily depends on the correspondences of these points [59]. The SPHARM representation contains shape information and features of a 3D object. These 3D shape features allow us to estimate the orientation by evaluating the correlations (differences) between two 3D objects.

Let's say we have a 3D objects template, W and the other 3D object, P, that shares similar shape but with different orientation. We first compute the SPHARM estimation of W and P and get two sets of SPHARM coefficients w and p for the two objects respectively. We want to find a rotation matrix R_{3x3} that estimates the rotation of these two sets of SPHARM coefficients that minimizes the difference between two sets of coefficients, w and p. The difference can be formulated as

$$\sum_{k=1}^{l} \|Rp_k - w_k\|^2, \tag{3.6}$$

where $w_k = w_1, w_2, \dots, w_l$, $where\ l$ is the number of coefficients. w_k is the coefficient set of the template SPHARM model and $p_k = p_1, p_2, \dots, p_l$ is the coefficient set of the nuclei SPHARM model. The equation (3.6) can be expanded as

$$\sum_{k=1}^{N_b} \|Rp_k - w_k\|^2 = \sum_{k=1}^{N_b} (Rp_k \cdot Rp_k) - 2 \sum_{k=1}^{N_b} (Rp_k \cdot w_k) + \sum_{k=1}^{N_b} (w_k \cdot w_k) \tag{3.7}$$

$$= \sum_{k=1}^{N_b} (\|p_k\|^2 + \|w_k\|^2) - 2 \sum_{k=1}^{N_b} (Rp_k \cdot w_k). \tag{3.8}$$

Since the coefficients p_k and w_k are considered as constant during the alignment process, the first summand in (3.8) does not change along with the rotation matrix R. Therefore, to minimize the difference of two objects in equation (3.7) is equivalent to maximizing the second summand in equation (3.8), as

$$max \sum_{k=1}^{N_b} (Rp_k \cdot w_k). \tag{3.9}$$

Here we implement a rotation estimation method base on SPHARM descriptor introduced by Althloothi *et al*.[59]. The method is to find a unit quaternion vector q with a symmetric 4 x 4 matrix M such that

$$\max_{R} \sum_{k=1}^{N_b} (Rp_k \cdot w_k) = q^T M q, \tag{3.10}$$

where the matrix M derives from the SPHARM coefficients of template P and object W. It is calculated with cross-covariance matrix between the SPHARM coefficients of two objects as described in [59]. The unit quaternion vector q that maximizes (3.9) can be solved by using eigen-decomposition technique with matrix M. It has been proved that the quaternion vector q which satisfies (3.10) is the eigenvector that corresponds to the maximum eigenvalue of matrix M. Thus, the optimal rotation matrix, R, that aligns a 3D object to a 3D template can be computed by converting the unit quaternion vector q. After all the nuclei were aligned to the template shape, the CTs within them also were rotated to the common coordinate by applying the rotation matrix of each individual nucleus to the CTs within the corresponding nucleus as shown in Figure 3.13.

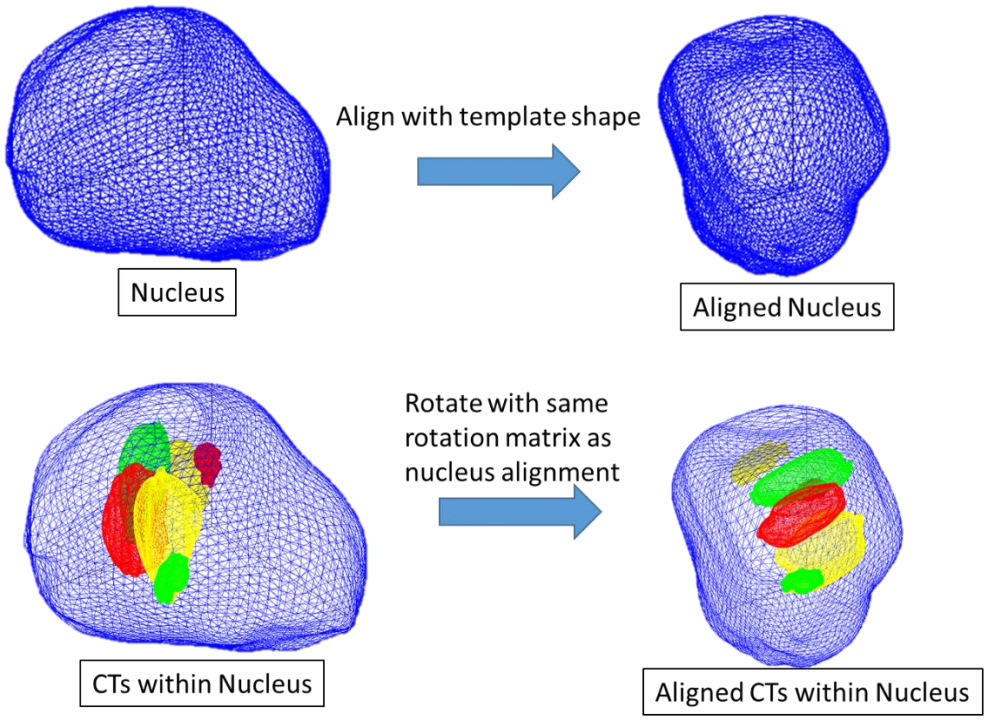

Figure 3.13. Alignment based on nucleus shape applied to CTs.

Nuclei were aligned to a template SPHARM model generated by finding a shape that has the minimum sum of RMSD to corresponding SPHARM coefficients of all nuclei. The template is defined by computing the average shape of a set of nucleus samples. The SPHARM-MAT toolbox provides an utility that is able to find an average shape given a set of objects in terms of SPHARM shape descriptor [55]. The average function includes the following steps: (1) Randomly select one of the sample as initial template, (2) aligning all other samples to the initial template (in frequency space without altering the shape in the object space) [60], and (3) finding a set of SPHARM coefficients that has the minimal RMSE to the SPHARM coefficients from all the nucleus samples. The average shape is shown in Figure 3.14.

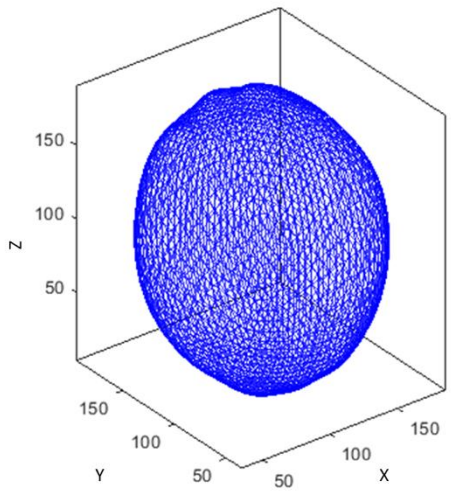

Figure 3.14. The average shape as alignment template.

3.5.2 Selection of Structural Landmarks and Alignment

After aligning all nuclei and CTs within them to the template shape using SPHARM coefficients as described above, the 3D positioning of CTs in 3D space within a nucleus was determined by their spatial position respect to a structural landmark. The structural landmark was selected to be one of the chromosomes from the FISH image.

Homologs in each class were categorized into specific groups based on the level of ploidy. For example, each homolog from the pair of diploid X chromosomes may be categorized as belonging to the group X-1 or X-2. This is critical since most of the chromosome exist in pairs in the normal cell, and some may express differently in an abnormal cell. In this study, we evaluated the CT's of chromosome X, 8, and 12. For a normal cell, it has two of each chromosome. On the other hand, an abnormal cell has two of chromosome X and 8 as well as three of chromosome 12 (trisomy 12). We need to

specify which particular homolog chromosome that we are referring to for further analysis of the 3D position.

The categorization was performed based on the assumption that the CT of a given chromosome class occupies a similar position in 3D space for a given population. The position of a CT was represented by a vector originating from the origin of the common reference frame (i.e., centroid of the nucleus) to the centroid of the CT. One class from normal chromosome class (i.e., chromosome X or 8) was selected as reference class. We then picked one nucleus sample as the reference sample then choose one homolog from the reference class to be the predefined global structural landmark. The homologs that belong to the same class as the global landmark were assigned to specific group by computing a scalar value, i.e., the vector dot product, between their position and the global structural landmark. A grater magnitude of the scalar value indicated a closer proximity to the global structural landmark. The homologs from each sample that was closer to the global landmark were assigned to group 1 and the further ones are group 2. The group 1 homologs in the class are assigned as the local structural landmark in each sample. The categorization of the homologs in the other classes are done according to their proximity to the local structural landmark in each nucleus. Based on this value each chromosome was assigned to the appropriate group (e.g., 1 or 2 for diploid and 1, 2 or 3 for triploid chromosomes). For example, if we select chromosome X as the reference class, homologs of chromosome X in each sample are categorized to group X-1 and group X-2 according their proximity to the global structural landmark. Then, for each sample, compute the dot product of homologs from other chromosome classes and the chromosome X-1 in the nucleus and categorize accordingly.

All the local structural landmark of nucleus samples are aligned to the position of the global landmark from the reference nucleus. This is done by rotating all the sample nuclei such that the local landmarks and the global landmark are along the same radial line. The rotation transformation is done by three parameters (angles) ϕ, θ, and ψ that represent the angle rotated along z-, y-, and z-axis as shown in Figure 3.15.

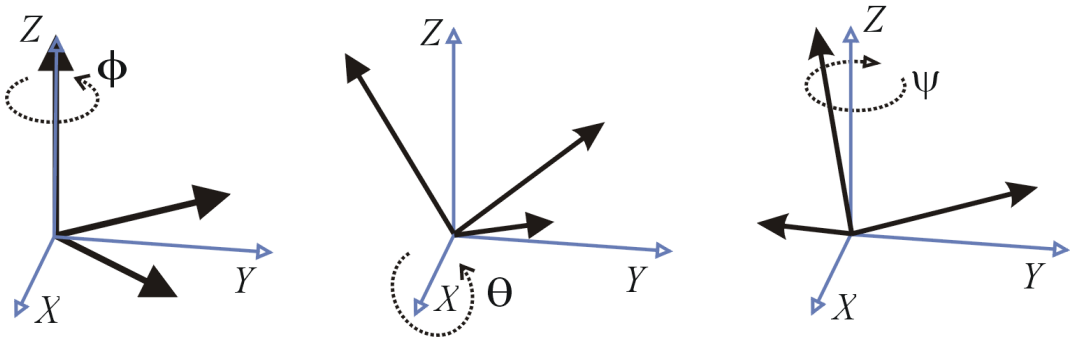

Figure 3.15. Euler angles of three rotation operations.

The rotational transformation is conducted with rotational matrix T,

$$T = \begin{bmatrix} n_x & O_x & a_x \\ n_y & O_y & a_y \\ n_z & O_z & a_z \end{bmatrix}, \tag{3.11}$$

where

$n_x = \cos(\emptyset) * \cos(\theta);$

$n_y = \sin(\emptyset) * \cos(\theta);$

$n_z = -\sin(\theta);$

$O_x = \cos(\emptyset) * \sin(\theta) * \sin(\varphi) - \sin(\emptyset) * \cos(\varphi);$

$O_y = \sin(\emptyset) * \sin(\theta) * \sin(\varphi) - \cos(\emptyset) * \cos(\theta);$

$O_z = \cos(\theta) * \sin(\varphi);$

$a_x = \cos(\emptyset) * \sin(\theta) * \cos(\varphi) + \sin(\emptyset) * \sin(\varphi);$

$a_y = \sin(\emptyset) * \sin(\theta) * \cos(\varphi) - \cos(\emptyset) * \sin(\varphi);$

$a_z = \cos(\theta) * \sin(\varphi)$.

This rotation matrix is defined by multiplying the individual rotation matrices corresponding to the rotation operations in Figure 3.15.

3.5.3 Scaling local landmark respect to global landmark

After aligning the local landmark of samples with global landmark, all the local landmarks in each sample is at the same radial direction, but may lie at different radial positions (magnitude) depending on the normalized distance. In order to find a common magnitude, the position of all chromosome in each sample is scaled with a scaling factor. The scaling factor can be different from sample to sample and each individual scaling factor is computed as the ratio of the magnitude of local landmark to the global landmark from the reference cell given as:

$$scaling\ factor = \frac{Normalized\ magnitude\ of\ global\ landmark\ vector}{Normalized\ magnitude\ of\ local\ landmark\ vector}. \quad (3.12)$$

The spatial position of all the chromosomes within each nucleus is evaluated by computing the relative positioning to the local landmark and scaling with the scaling factor from (3.12). The alignment and scaling procedure allow the comparison of all chromosomes in different nucleus samples with a common 3D benchmark position.

3.6 Quantify Position and CT proximity Relationships

In the sections above, we describe how the CTs and nuclei are aligned to a common coordinate system that determines the relative position of each chromosome homolog in terms of specific (x,y,z) coordinate. This allows a comparison of the spatial organization of CTs across multiple nuclei from a given population and/or between populations.

3.6.1 Radial and Edge to Edge distance

CT positioning is quantified using two different types of relative distances: radial and edge to edge distance. The radial distance is the Euclidean distance between the centroid of the enclosed nucleus and the centroid of each chromosome within the nucleus. The edge to edge distance of two 3D objects is defined by the shortest Euclidean distance between the surfaces of the two objects along the line that links the centroid of the two objects. The ray triangle intersection method was used to determine the points where the line segment passing through the centroids intersects the objects surface[61]. Figure 3.16 presents a schematic of the radial distance (Gc) and edge to edge distance (GRs) of two objects.

The radial distance is normalized by the Euclidean distance of intersecting point to the center of the nucleus (Gf). The normalized edge-to-edge distance between two signals is computed as the ratio GRs/GRf as shown in Figure 3.16.

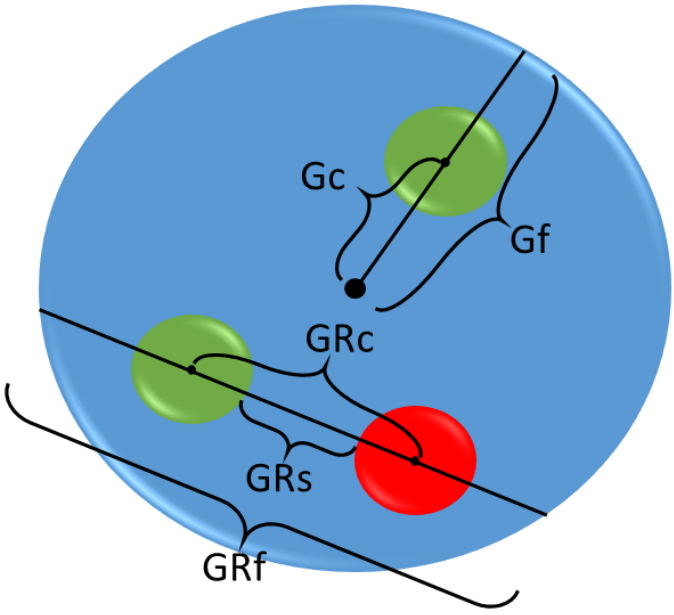

Figure 3.16 Schematic of radial and edge to edge distances.

3.6.2 Characterization of Intra-homologous and Inter-heterologous Distances

To fully assess the relative location of each chromosome, the distance between each individual chromosome within nucleus was determined. This information can be critical for biological processes. For example, neighboring chromosomes may carry co-regulated genes. The intra-homologous distance is the distance between homologous chromosomes of the same class (e.g., the distance between two X chromosomes (X-X)) and it is defined as the edge to edge distance between them. The inter-heterologous distance is the edge to edge distance between different chromosomes (e.g., distance between chromosome X and chromosome 12 (X-12)).

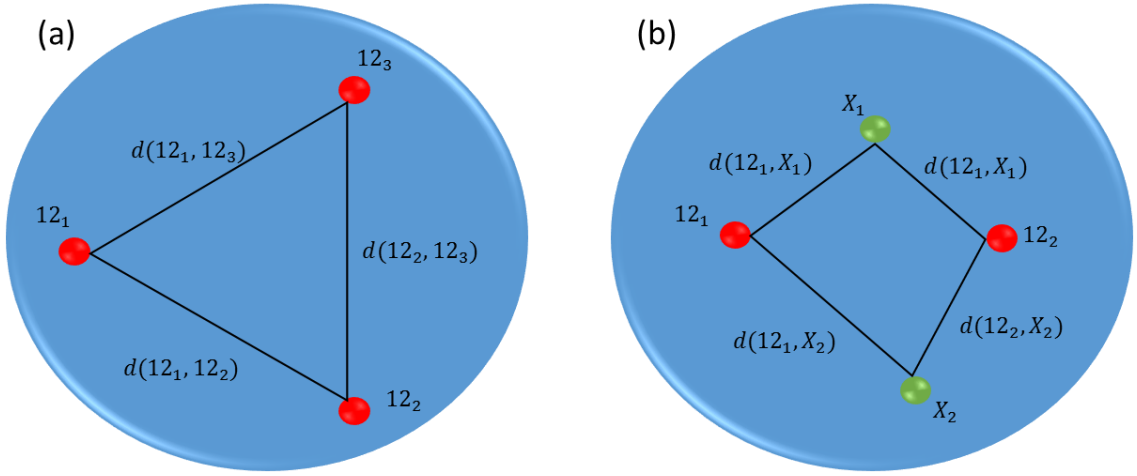

Figure 3.17. (a) Intra-homologous distance of a trisomy 12 and (b) inter-heterologous distance of chromosome X and chromosome 12.

The normal cells are diploid and only have one intra-homologous distance for each chromosome within the nucleus. In case of aneuploid cells, we have trisomy 12 for this study, the intra-homologous distance can have three values as shown in Figure 3.17 (a). The minimum of the three distances is selected to be the intra-homologous distance for a given nucleus such that

$$d_{hom} = min\big(d(12_1, 12_2), d(12_1, 12_3), d(12_2, 12_3)\big). \quad (3.13)$$

There can be several inter-heterologous distances depending on how many chromosomes are in the nucleus. For example, for a normal cell, the inter-heterologous distance of chromosome 12 and X have four distance values: $d(12_1, X_1), d(12_1, X_2), d(12_2, X_1), and\ d(12_2, X_2)$ as shown in Figure 3.17 (b). We calculate the mean of the minimum inter-heterologous distances to represent the heterologous distance between the two chromosomes. As shown in Figure 3.17 (b), the heterologous distance of chromosome 12 and chromosome X in a normal cell can be formulated as

$$d_{het} = \frac{[min(d(12_1,X_1),d(12_1,X_2)) + min(d(12_2,X_1),d(12_2,X_2))]}{2}. \quad (3.14)$$

3.6.3 Statistical assessment of the 3D position

Statistical analysis provides an objective insight/overview for a given data set or population. Here we perform student's t-test for comparing the position (centroid) differences between diploid versus aneuploid samples. In order to conduct the statistical assessment on the position, the 3D position in (x,y,z) coordinate is transformed to a single scalar value. This is achieved by computing the scalar projection of the centroid of CTs in direction of the local landmark within nucleus. Figure 3.18 shows the scalar project of a vector of CT centroid in direction of vector of the landmark is defined as $|Vector_{CT}| * cos\theta$, where θ is the angle between vector of CT centroid and vector of the landmark.

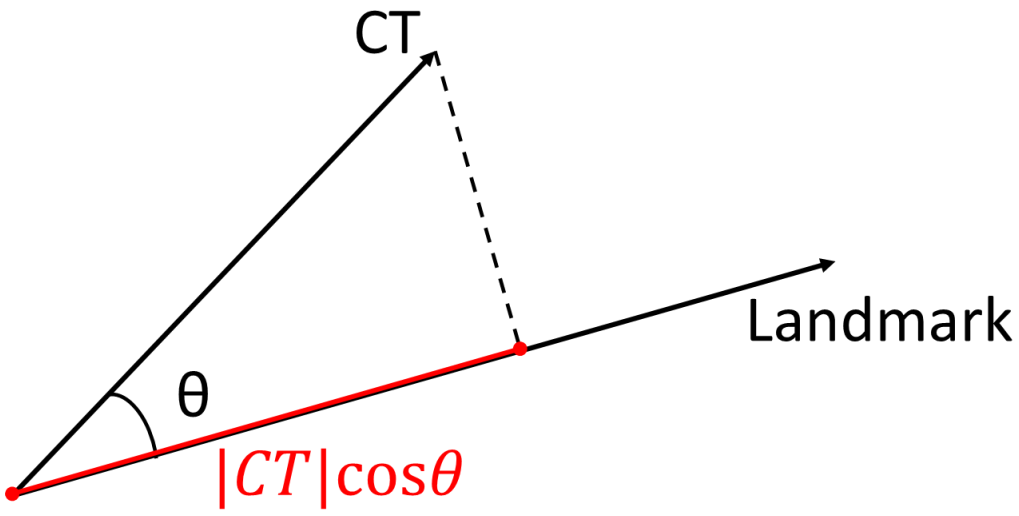

Figure 3.18. Projection scalar value of CT vector in direction of vector of the landmark

3.6.4 Nuclear and CT Volume

The volume of each chromosome territory and nucleus are calculated by finding the volume enclosed within the reconstructed 3D mesh surface. In order to compute the volume inside the mesh object, divergence theorem or Gauss's theorem is used with the vertices and faces of the object. The divergence theorem states that the outward flux of an enclosed surface equals to the volume integral of the divergence over the region within the enclosed surface. It can be written in the form:

$$\iint_S \vec{F} \cdot d\vec{S} = \iiint_V div(\vec{F}) dV, \tag{3.15}$$

where V is a solid region of $\mathbb{R}^n$ (here n = 3 such that V represents a volume in 3D space), S is the boundary surface of V, and $\vec{F}$ is a vector field that can be continuously differentiable in the neighborhood of V. The $div(\vec{F})$ represents the divergence of vector field $\vec{F}$ which can be formulated as

$$div\big(F(x,y,z)\big) = \frac{\partial F(x,y,z)}{\partial x} + \frac{\partial F(x,y,z)}{\partial y} + \frac{\partial F(x,y,z)}{\partial z}. \tag{3.16}$$

Given the SPHARM reconstructed surface (triangular mesh) and norms, we are able to compute the surface area of S and the divergence and then find the volume of each CT and nucleus object.

Chapter 4 – Validation

Validation was performed to assess the integrity of proposed 3D modeling system, including the algorithm for image thresholding, SPHARM modeling, and alignment.

4.1 Image segmentation

The most important part of image segmentation for this study is to select a threshold value that performs properly and separates the object from background. To evaluate the performance of the threshold value, the gray-level co-occurrence matrix (GLCM) using the eight-connected pixel neighborhood was computed for each slice of the image stacks. A co-occurrence matrix gives the spatial relationship of pixels for a given image. It measures the frequency at which pairs of pixels with specific values occur in an image. The size of co-occurrence matrix depends on the dynamic range of the image and the values in the matrix are computed by counting how often two specific values appear together. Compared to a histogram, a co-occurrence matrix does not only give the pixel value distribution of the image, but also provides its spatial information. Figure 4.1, gives an example of how to compute a co-occurrence matrix, M, for a given gray scale image, I. We examine all the pixels in a gray scale image, find its value and count the occurrence of certain value according to its offset (i.e., the distance between pixel of interest and its neighbor). The rows of co-occurrence matrix represent the values of the pixel that is currently examined (pixel of interest) and the columns refer to the values of its neighbor pixel. In the example in Figure 4.1 the offset is the pixel on the right of the pixel of interest. When the value of current pixel is 0, its neighbors with a value of 1 only appear once; thus, the value for co-

occurrence matrix $M(1,2) = 1$ (as seen in the red boxes). As for all pixels in I that equal to 1, two of them have neighbor with value of 3. Therefore, $M(2,4) = 2$.

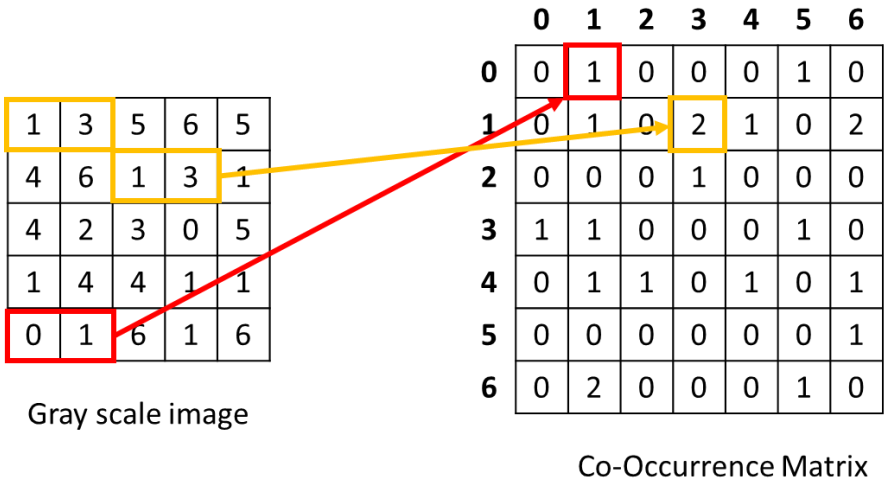

Figure 4.1. Defining a co-occurrence matrix for a given gray scale image.

Our images of chromosomes and nuclei can be characterized into three regions: (1) background, (2) blur noise region, and (3) the object as shown in Figure 4.2 (A). Manual inspection was used to identify these characteristic regions in the corresponding co-occurrence matrix (shown in Figure 4.2 (B)).We define the three regions as follows: background region in the image with lower intensities would cluster at the top-left corner of the co-occurrence matrix, while object regions with much higher intensity values would cluster at the bottom-left of the co-occurrence matrix. The blur region around the imaged CT's has a wider value range and appears along the diagonal line of the matrix with less condensed from. The goal here is find a threshold value that separates region (2) and region (3). The threshold value should locate on the diagonal line of co-occurrence matrix between the two characteristic regions relative to region (2) and region (3) on the input image. Figure 4.2 (B) gives an example of proper threshold value (the red asterisk in Figure 4.2 (B)) for segmenting image in figure 4.2 (A). 110 images were randomly selected from our

data set and evaluated using co-occurrence matrix. Out of 110 images, the threshold value of 94 images falls in desired region (between region 2 and region 3 as seen in Figure 4.2 (B)). Threshold values from the remaining 16 images are unable to separate the object from the background mainly due to the inadequate image quality. Figure 4.2 (D-F) gives an example of an image where the global threshold fails. Since the object is fairly blurry and the GLCM does not show any obvious characteristic regions, it is hard to determine a threshold value even for manual selection. We discard these images where the threshold fails and we cannot validate the detected objects. Out of 62 optical sections that contain all four channels, 56 samples images are useable including 27 diploid and 29 aneuploidy.

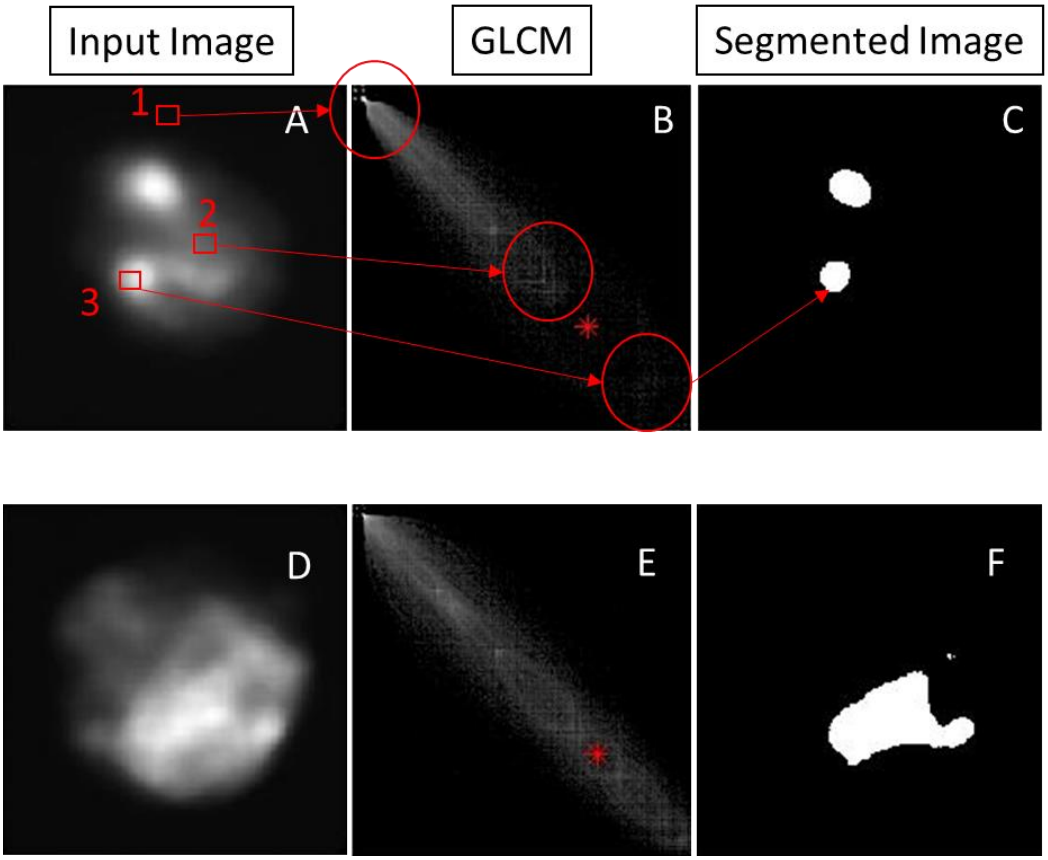

Figure 4.2. Gray-level co-occurrence matrix (B and E) derive from gray scale image (A and D) and segmented image (C and F).

4.2 SPHARM Modelling Accuracy

To test the accuracy of the modeling result, simulated shapes were used for the evaluation. We created four volumetric objects: a sphere, an ellipsoid, ellipsoid with rotation, and a deformed ellipsoid as shown in Figure 3.9. RMSE was used to evaluate the performance of the SHPARM modeling. Results are shown in Table 4.1. For all test objects, the RMSE was less than 0.03 indicating a mean error of 0.0256 with a standard deviation of 0.0036 in using SPHARM modeling.

Table 4.1. Accuracy evaluation of four simulated objects.

Test Object	Sphere	ellipsoid	Rotated ellipsoid	Deformed ellipsoid
RMSE	0.0286	0.0214	0.0237	0.0286

4.3 SPHARM shape descriptor based nuclei alignment

This section assesses the robustness of nuclei alignment that is based on shape information (i.e., SPHARM coefficients). We first test the reliability of the method by using simulated objects, followed by an evaluation with our experimental data set.

4.3.1 Alignment test

To test the reliability of the shape-based alignment algorithm, we select one nucleus sample from our data set then modify its shape to generate multiple nuclei simulations of similar but not identical shape, by scaling each SPHARM coefficient by a random number between 1 and 2 (normal distribution with mean = 1.5 and standard deviation = 0.1) (as shown in Figure 4.3). Next the modified nuclei are rotated to different testing angles along x-, y, and z-axis, and aligned to the original sample and the recovery angle is computed.

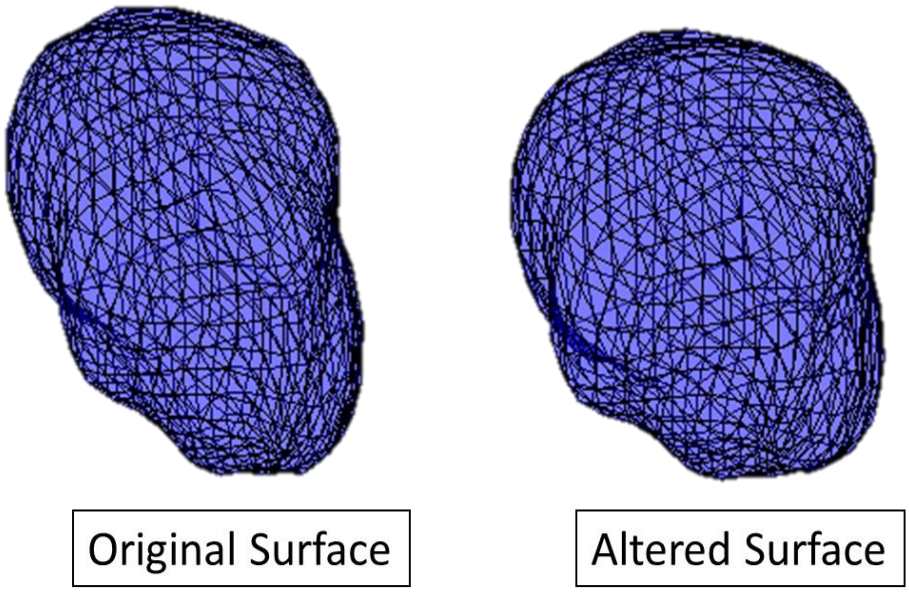

Figure 4.3. A test nucleus surface and its altered surface.

The test objects were defined by rotating the altered object along x-, y-, and z-axis at degrees in 15, 45, and 75 then we have a total number of 27 different rotation angles (e.g., [15 45 75], [15 75 75], and [45 75 15] etc.). Each test object at different angle is aligned with the original surface using the method we described in (3.5.1 The method is to find a rotation matrix that rotates the test object so that the RMSE of coefficients between test object and the template (i.e., the original surface here) is minimized. The angular error is defined by the angular difference between the true rotation and the estimated rotation angle along x-, y- and z- axis, and is used to evaluate the performance of the alignment. We also computed the angular angle between the test objects and altered object.

Table 4.2. Alignment test using a sample nucleus and its altered shape.

Alignment template	X-axis difference	Y-axis difference	Z-axis difference
Angular error between Altered surface and Altered rotated object			
Mean	3.86E-13	9.02E-14	4.67E-13
Standard deviation	4.15E-13	8.41E-14	4.83E-13
Angular error between Altered rotated object and the original object			
Mean	12.64426	2.903537	9.698851
Standard deviation	8.508068	1.869545	9.723114

Table 4.2 shows the angular error along x-, y-, and z- axis when the test objects are aligned to the original object and altered object. As the test objects are aligned to the altered surface, the angular error along three axis are very close to zero which proves that this method is able to align two objects that have identical shape but are at different orientations. The angular error when aligning the altered and rotated objects to the original surface is relatively high. This error is caused by the shape differences between the two surfaces. To visually observe the performance of the alignment, some test objects before and after the aligning to the original object are shown in Figure 4.4. The visualization shows that regardless the initial orientation of the test objects, all of them are rotated to similar orientation after aligned to the same template. As a result, the first step of alignment process is able to provide alignment of all the nuclei under the assumption that nuclei share similarity in shape.

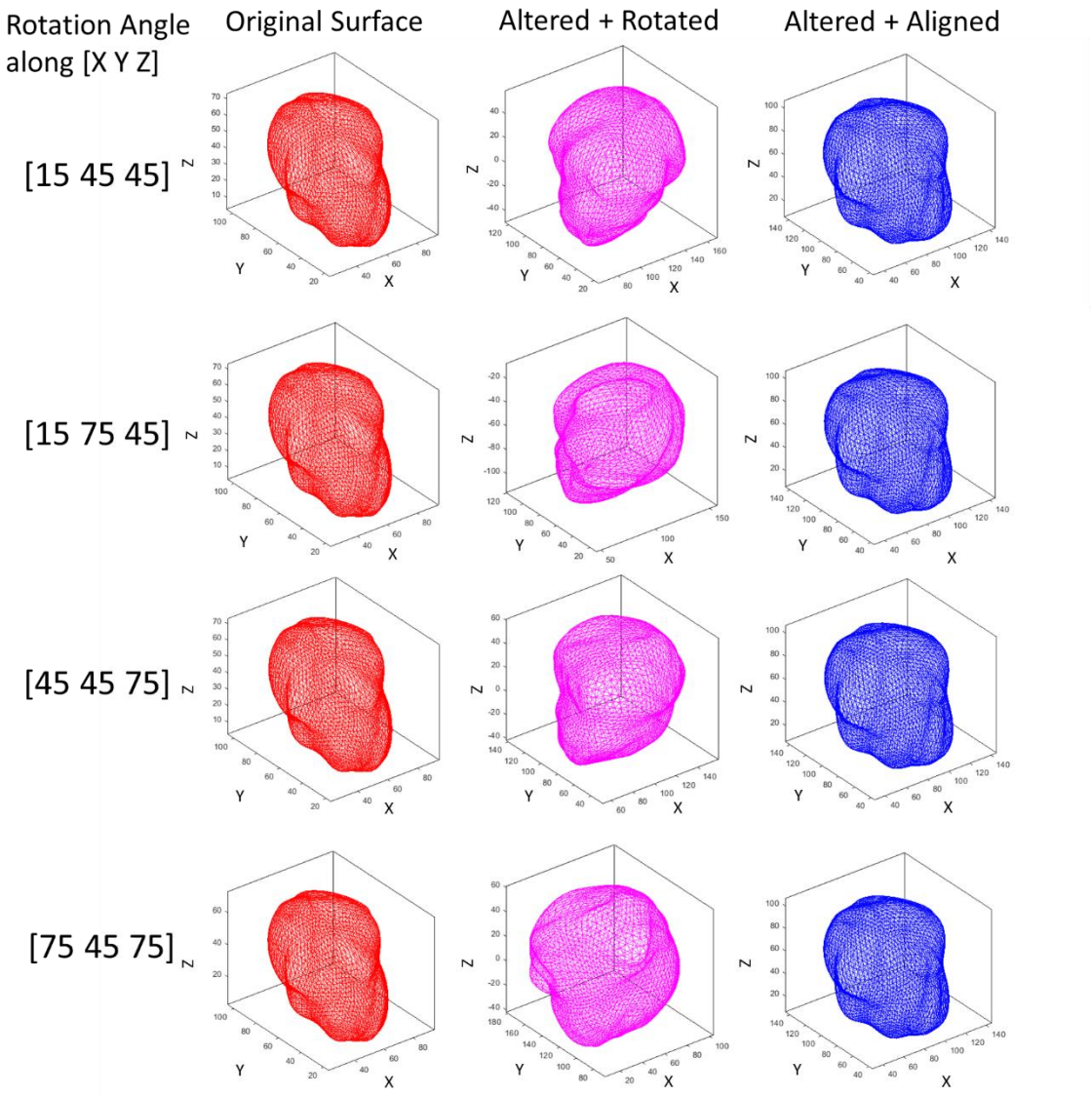

Figure 4.4. Examples of test objects aligned to the template object (original surface).

4.3.1 Template Selection and Alignment

The performance of the modeling framework relies upon the template utilized for aligning all nuclei in a given population. In this section we assess the alternative of using a single nucleus as a template versus using the average shape of all nuclei as the shape template.

We have a total of 56 nuclei samples for the study and we would like to see how the result of our system would be affected when using each one of the nuclei as a template. To do so, we use each one of the nuclei samples as the template and have all the other nuclei aligned to it.

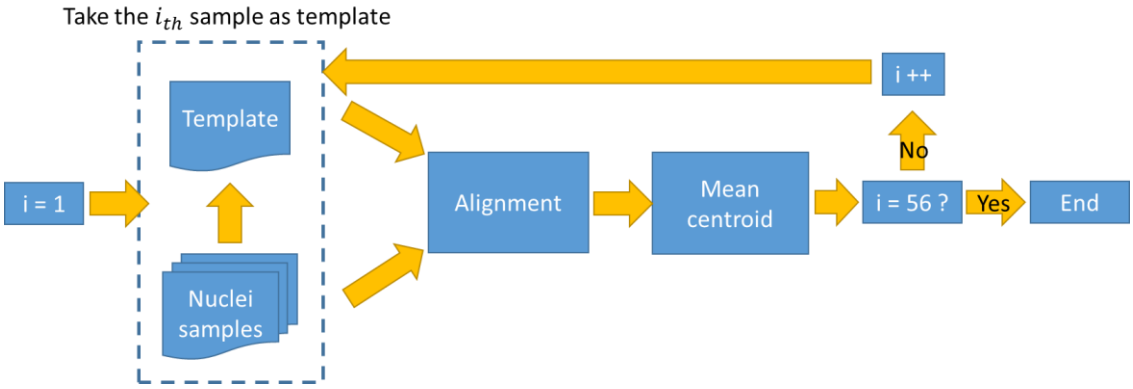

Figure 4.5. Process of shape based alignment validation with different templates.

Figure 4.5 shows the validation process. We select the i_{th} (i = 1 to 56) nucleus from our nuclei sample set as the template to perform the shape-based alignment and landmark selection and modeling. The process repeats 56 times until we have used all nuclei as the template. To compare the position of chromosomes after the alignment, the mean centroid of each chromosome aligned with the template is calculated for evaluation. Figure 4.6 shows the mean position of chromosomes from aneuploid cells (a) and diploid cells (b) and using average shape as template (c and d). Different colors represent different chromosome and the shape of the marker shows the homologous group of a chromosome. Green, red, and blue represent chromosome X, chromosome 12, and chromosome 8 respectively. The square marker is the first homologous group, the asterisk is the second group, and circle is the third group. For example, the green asterisk represents chromosome X-2. This alignment uses chromosome X-1 (the green square in Figure 4.6) as landmark and thus the position of it is always the same across the 56 tests. The chromosomes that belong to the

same homologous group are clustered together in both diploid and aneuploid nuclei even with each different nucleus as the alignment template.

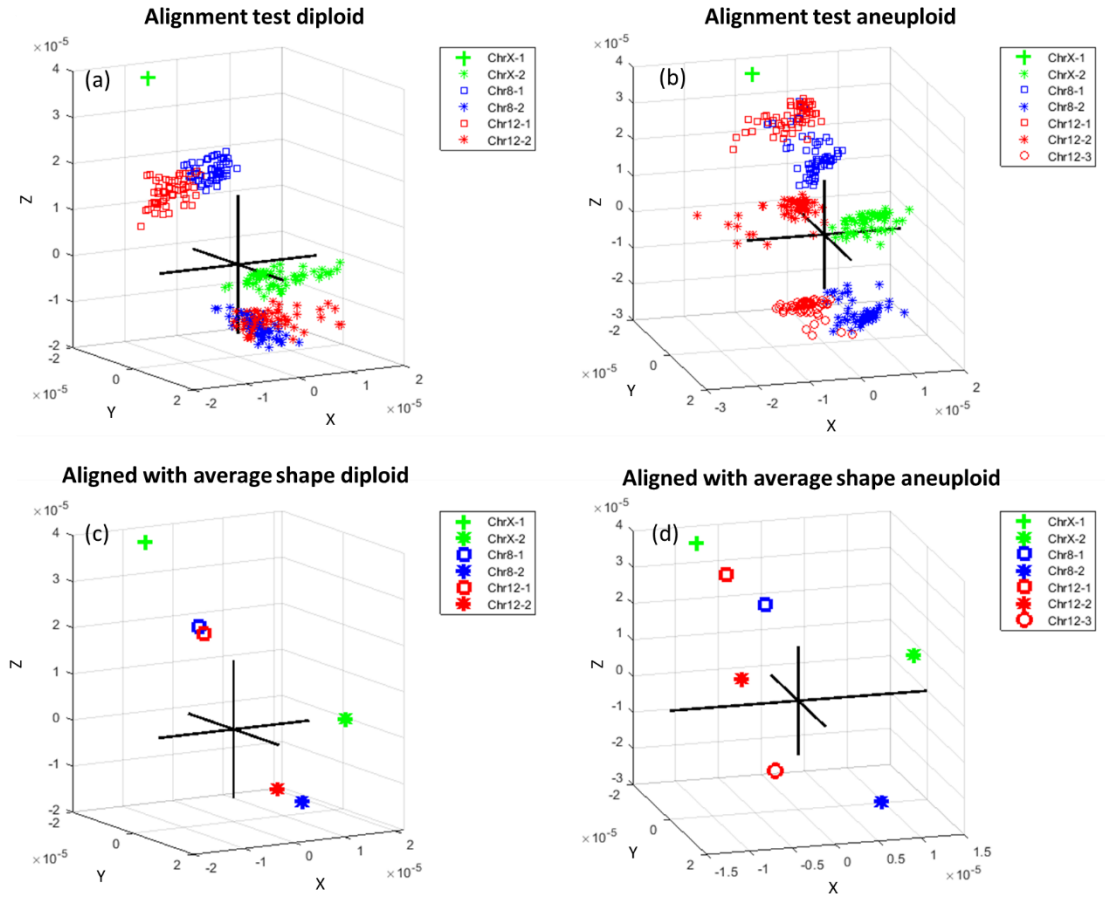

Figure 4.6. Mean centroid of chromosome plot from 56 alignment tests and the mean centroid of chromosomes using average shape as template.

4.4 System validation

The 3D surface of nuclei and chromosomes are reconstructed with SPHARM shape descriptor. The spatial organization of the chromosomes within a nucleus are with to its local coordinate system after the 3D surface modeling. Thus, a common reference frame and structural landmark are needed to provide the 3D position of the chromosomes within the 3D space in the nucleus for analysis across samples. The common reference frame is determined with a shape-based alignment using the shape information from SPHARM

shape descriptor (the coefficients) by aligning the nuclei into a common coordinate system followed by landmark alignment and scaling.

This section presents system validation using simulated data with the methods described above. Here we use two sets of simulated data: non-random and random spatial organization of chromosome territories. Figure 4.7 shows the process of generating simulated data sets. For the non-random simulated set, one nucleus is randomly selected from our data set as a template, and we generate three pairs of chromosome within the 3D space of the nucleus (representing chromosome X, 8 and 12). This template nucleus and the simulated chromosomes are rotated into six different orientations. The positions of the simulated chromosomes are also slightly altered to mimic our experimental sample (i.e., similar spatial organization, but not identical). For the second simulated data set, a random spatial organization of chromosome territories, is generated using the same nucleus template in six different orientations with randomly distributed chromosomes within it.

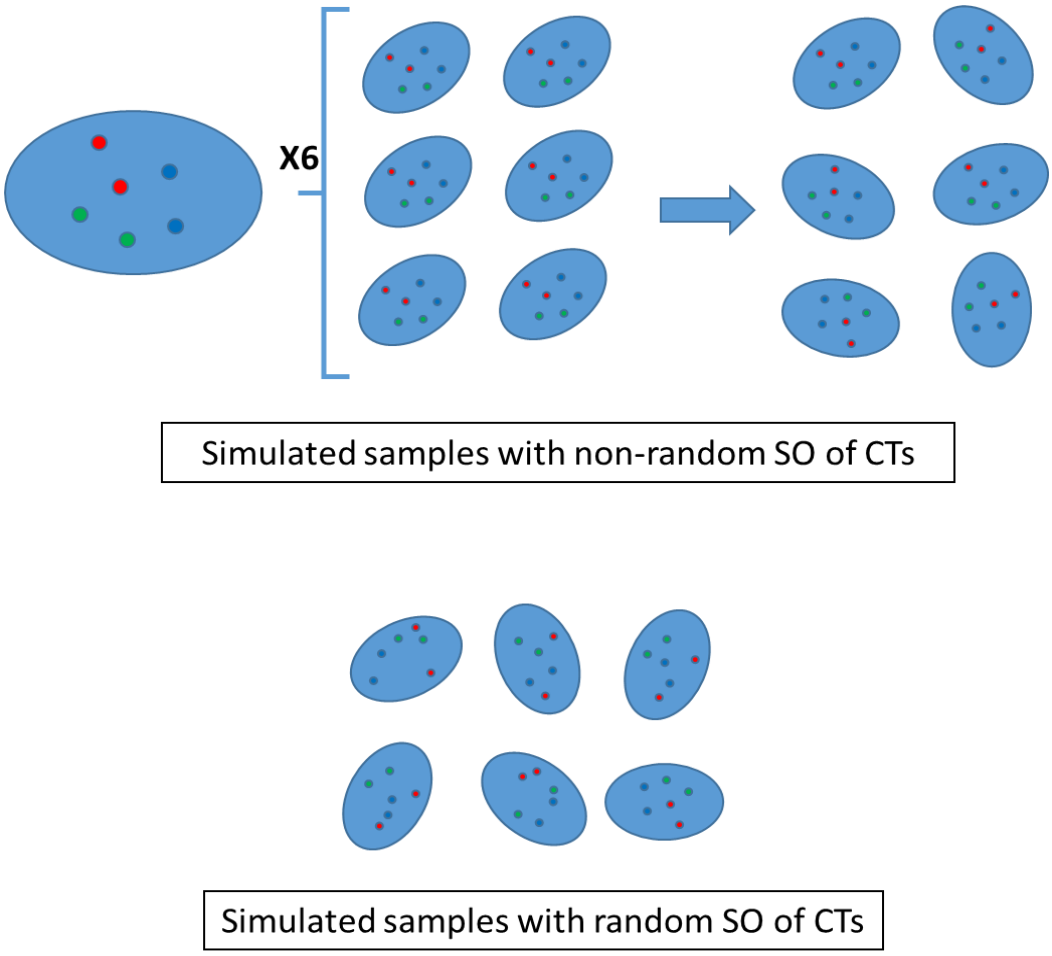

Figure 4.7. Simulated data sets for system validation.

Results from modeling the simulated sets are shown in Figure 4.8. The top one is the template nucleus and the colored dots within it are the six simulated chromosomes. Red and cyan channels represents chromosome 12 and 8 respectively. The black and green are chromosome X-1 group and X-2 group. For the validation, the black channel (chromosome X-1) is used as the local landmark for the structural landmark alignment and scaling. The remaining graphs show the six simulated nuclei in six different rotation angles and each of the color channels is located at similar corresponding positions to the template nucleus as described above.

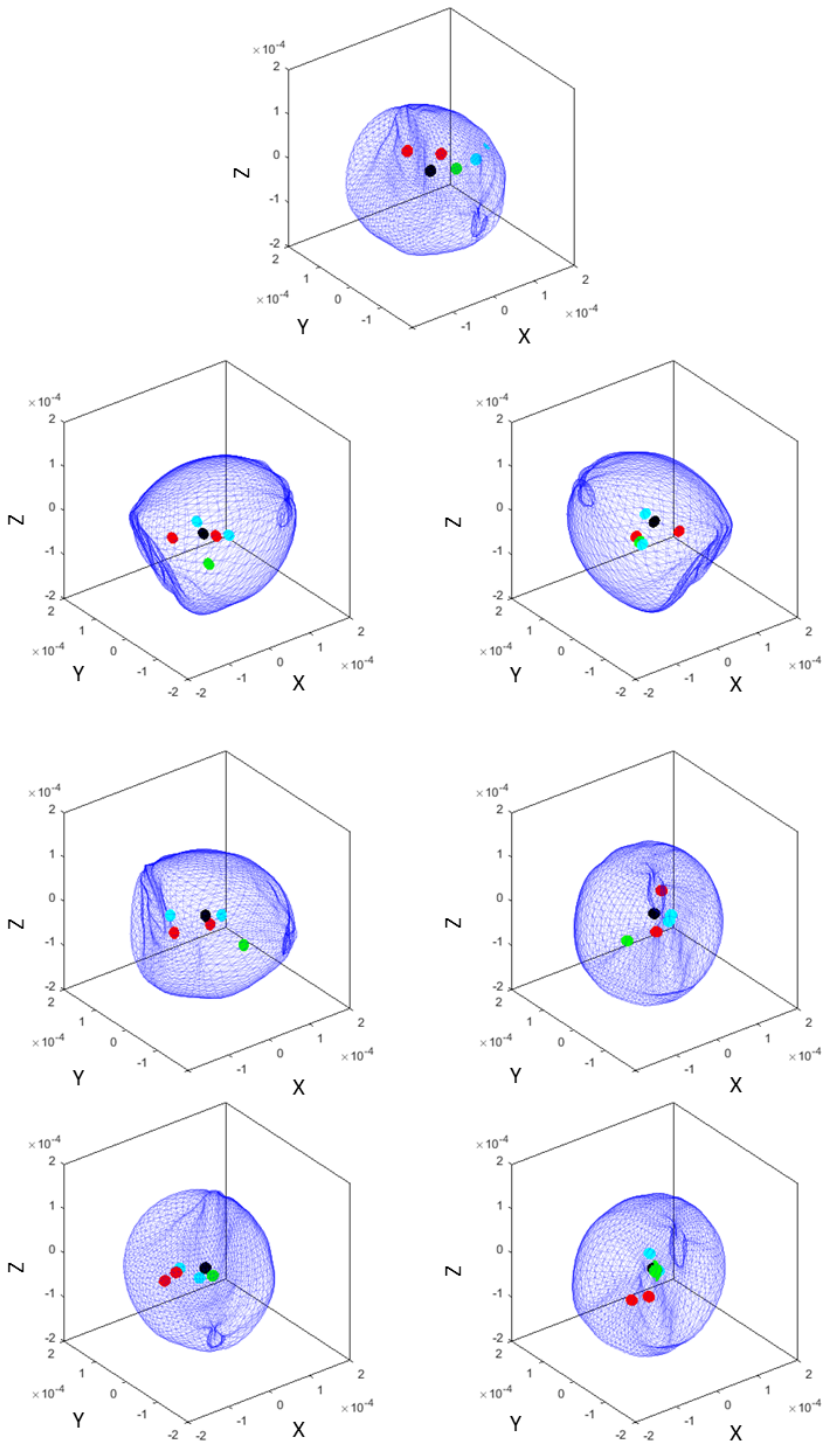

Figure 4.8. Simulated samples for nucleus that has chromosomes with non-random spatial organization.

To test the robustness of the modeling and alignment part of the system, the simulated data was generated as a binary volumetric image for the validation. The volumetric data of nuclei and chromosomes are transformed to triangular mesh surface and parameterized by mapping each vertices on the surfaces of a unit sphere. The coefficients of SPHARM shape descriptor are computed according to the vertices on the object and unit sphere. Finally, all the simulated data are aligned based on the shape information and structural landmark.

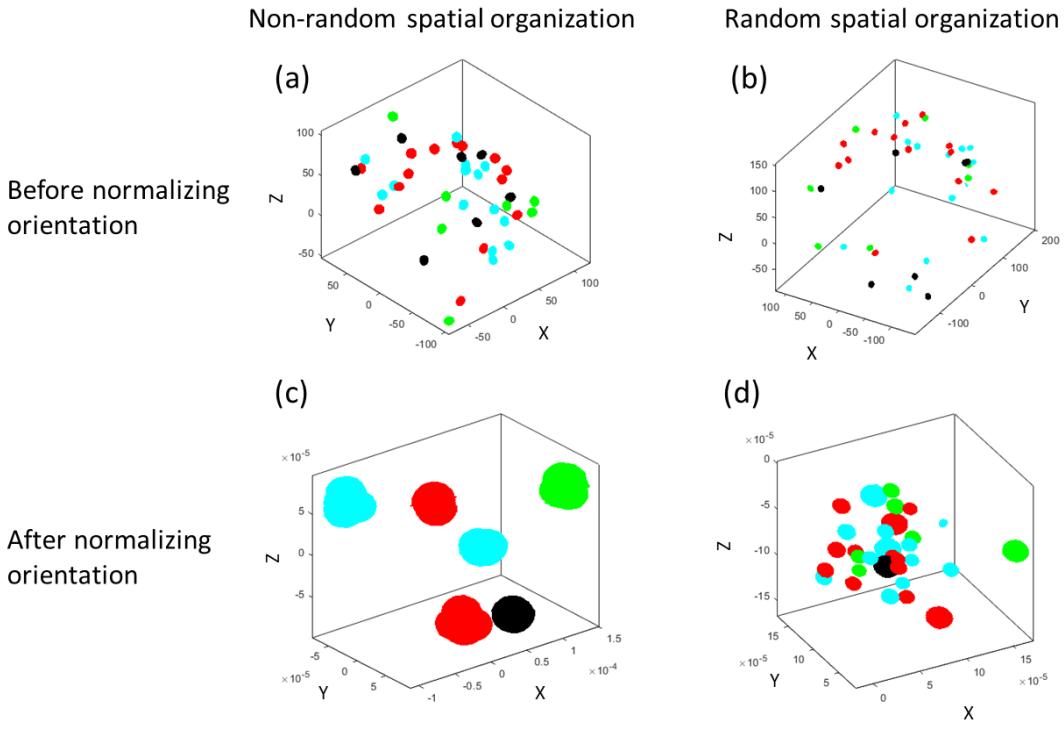

Figure 4.9. Chromosome localization of simulated data before (a and b) and after (c and d) the alignment.

These results validate the process of aligning nuclei samples to a common coordinate system for comparison across samples, including chromosome categorization, the use of a common reference frame, and shape based nuclei alignment, and structural landmark based determination of chromosome positioning. Figure 4.9 demonstrates the enclosed chromosome surface and position of simulated data in both random and non-random spatial

organization within the nuclei at different rotation angles. Prior to modeling, the localization of both random and non-random spatial organization data sets are not clustered as observed, due to the different orientations of the nucleus as shown in Figure 4.9 (a) and (b). Figure 4.9 (c) shows the chromosome localization of the non-random data set after modeling and alignment. The chromosomes that belong to the same homologous group are clustered within a small range and the structural landmarks (the black channel) of each nucleus sample are all located at the same position (global landmark). On the other hand, the random distributed group (Figure 4.9 (d)) shows no clustering except for the local landmarks that are aligned to the same position.

Chapter 5 – Experimental Results

Nuclei are sampled from a mixed population of diploid (~30%) and aneuploid cells (trisomy-12, ~70%) that are taken from cell cultures at passage number 73. The samples are processed with three-color FISH with the DAPI counterstained nucleus for visualization. The three-color FISH is targeting to chromosome X, 8, and 12 using whole chromosome painting probes with fluorescence dyes FITC (green), Aqua (cyan), and TRITC (red) respectively. A diploid normal cell presents two copies of chromosome X, 8, and 12. Each aneuploid cell has two copies of chromosome X and 8, and contains one extra copy of chromosome 12, i.e., three signals of chromosome 12.

5.1 3D Modeling using SPHARM

The FISH slides are prepared as described in section 3.2 The FISH signals are acquired using four channels (with laser wavelengths of 488, 458, 543, and 405 nm) confocal microscope at a z-interval of 0.3-μm and pixel size of 0.059 μm/pixel. The image slices are stored in form of 3D stacks of optical sections. The optical images of nuclei and CTs are preprocessed, identified according to region labeling, and segmented (as described in see 3.3). Each 3D object (including nuclei and CTs) is in form of binary 3D volumetric object and is transferred to triangular 3D mesh surface. Results are presented for a total of 27 diploid hES cells and 29 trisomy-12 hES cells. Figure 5.1 shows a representative 3D volumetric image (left) and its transformation into a triangular mesh surface (right). Each face is divided into two triangles by linking a diagonal line of the face and the vertices, and the edges on each voxel are used to compute the mesh surface.

The triangular mesh surface of nucleus and chromosomes, containing the vertices and face information, are parameterized by mapping the vertices on each surface in (x, y, z) coordinate onto spherical coordinate $(\theta, \emptyset)$ on a unit sphere. The mapping is done by minimizing the cost function that is designed for control of both area and length distortions.

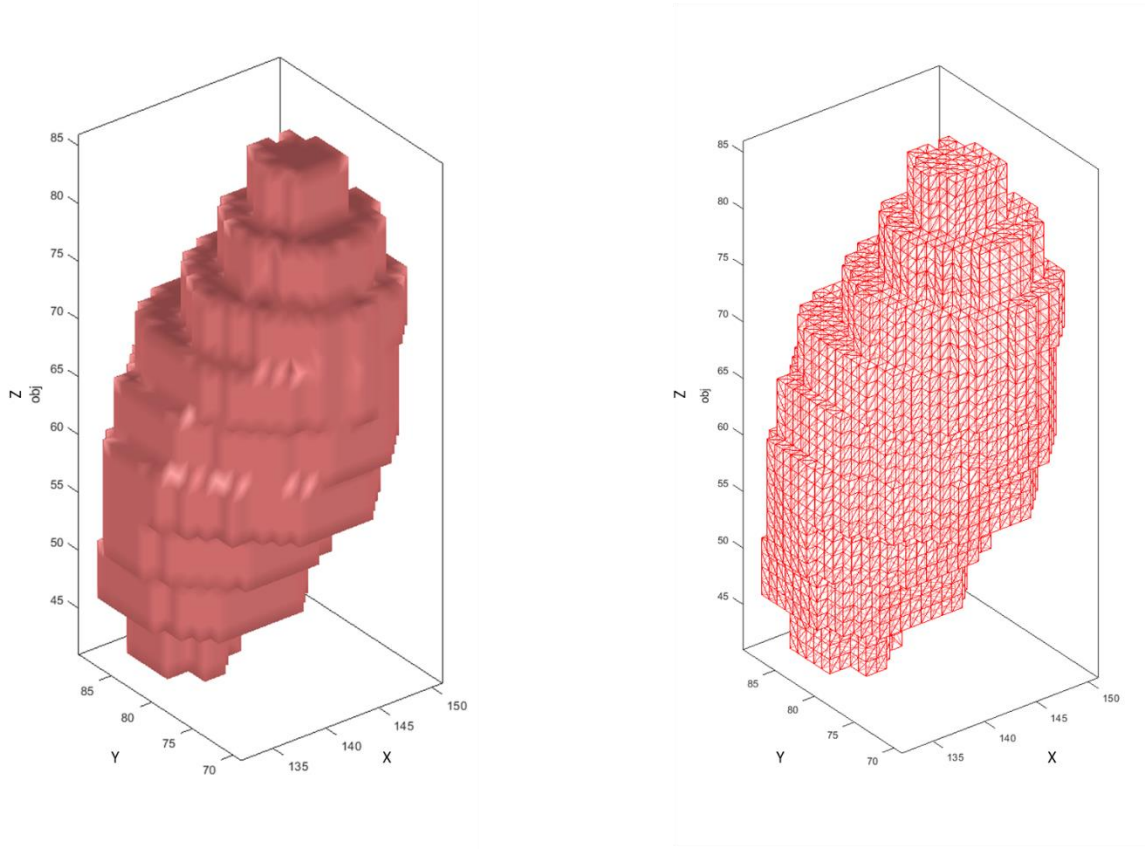

Figure 5.1. Volumetric data of a chromosome and its mesh surface after triangulation transformation.

As shown in Figure 5.2, the left figure is an example of the 3D triangular mesh surface of a chromosome and the figure on the right had side is the unit sphere and the corresponding vertices mapped from the triangular mesh. In both images, each vertex of the surface is color coded depending on their position and the faces values are interpolated

and colored accordingly. The same color gives the corresponding vertices on the input triangular mesh surface and the parametrized unit sphere.

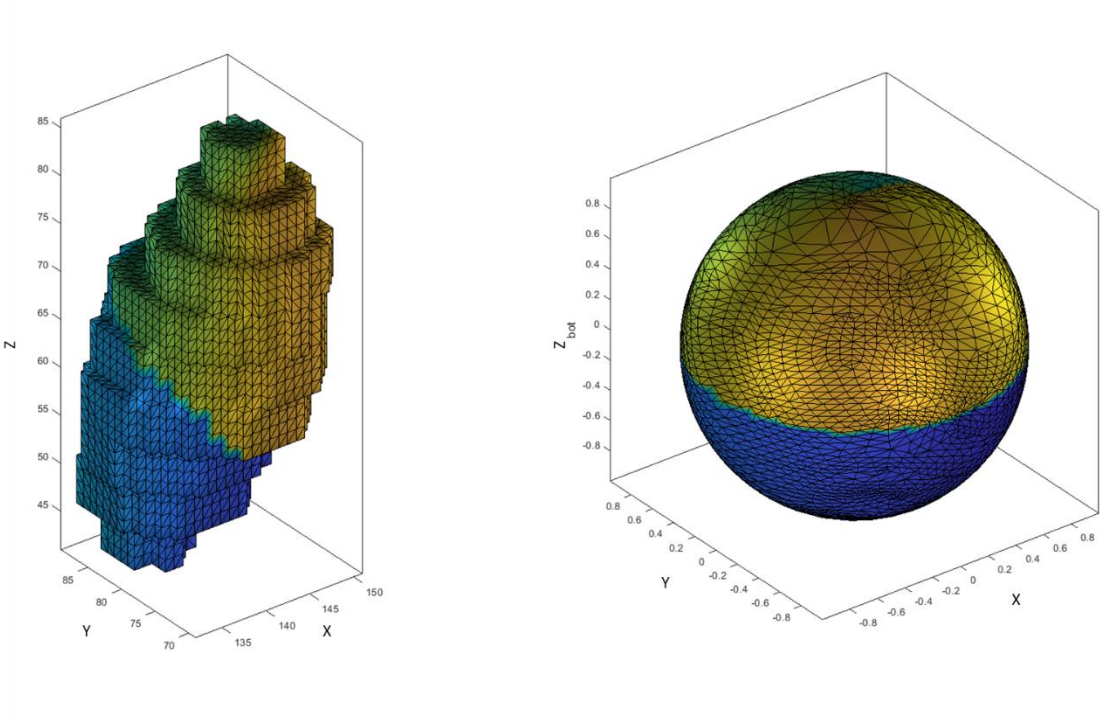

Figure 5.2. Spherical parametrization from (x, y, z) coordinate to spherical coordinate $(\theta, \emptyset)$.

The parameterized surface is next expanded into a set of spherical harmonic basis functions. Each basis function is multiplied with a scalar coefficient value that gives the magnitude of specific basis function at certain spherical location as described in section 3.4.2 The number of coefficients and basis function depends on the user defined degree. After the degree selection process described in section 3.4.2 this research uses a degree of 12 for the SPHARM shape descriptor for a surface reconstruction without expensive computation.

The set of coefficients of an enclosed surface contains the shape information in terms of magnitude of the SPHARM basis functions. A surface can be reconstructed with this

information and a set of sample points on a unit sphere. Here we use a set of sample points that are uniformly distributed on a unit sphere in spherical coordinate for shape reconstruction. The uniform sample on a unit sphere is pre-defined with a convex regular icosahedron with level 4 subdivision. Figure 5.3 shows a convex regular icosahedron (top) that has 30 edges and 20 equilateral triangle faces. Each triangle face is divided to four equilateral triangles at one subdivision. The bottom of the figure gives examples of level 1 to level 4 subdivision icosahedron. This study uses a level 4 subdivision icosahedron that contains 5120 triangular faces to reconstruct CTs and nuclei. Cartesian coordinate in (x, y, z) in the object space is computed with given spherical coordinates and SPHARM coefficients.

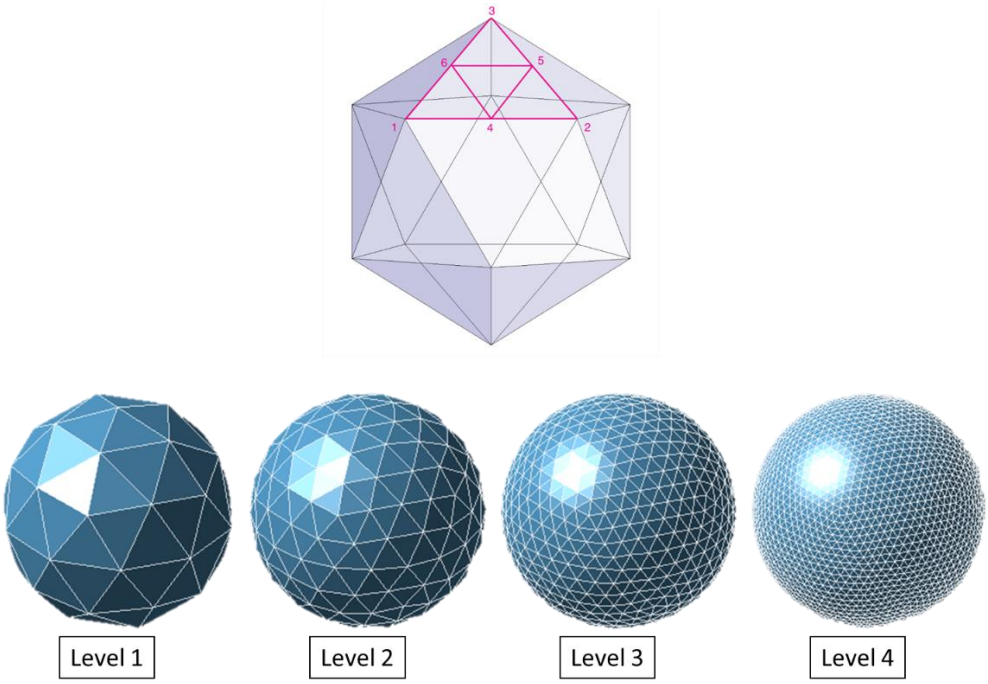

Figure 5.3. Convex regular icosahedron (top) and its subdivisions of level 1 to level 4 (bottom) [62].

The left figure in Figure 5.4 shows a mesh of the original vertices on a chromosome sample in green and the right figure is the reconstructed surface using SPHARM coefficients and the sampling vertices on a unit sphere. Figure 5.5 provides a visual comparison of the original object surface and the SPHARM descriptor reconstructed surface.

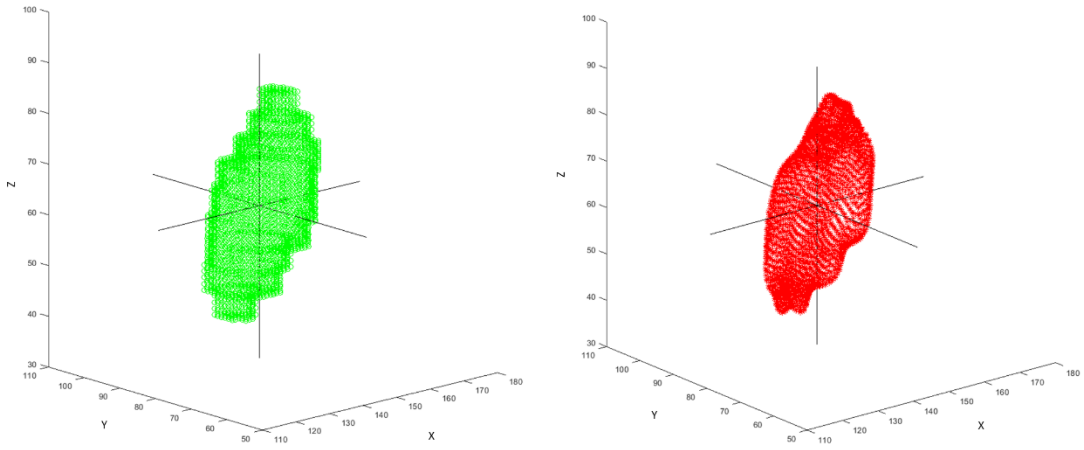

Figure 5.4. 3D surface of a chromosome sample object (green) and its reconstructed surface using SPHARM shape descriptor (red).

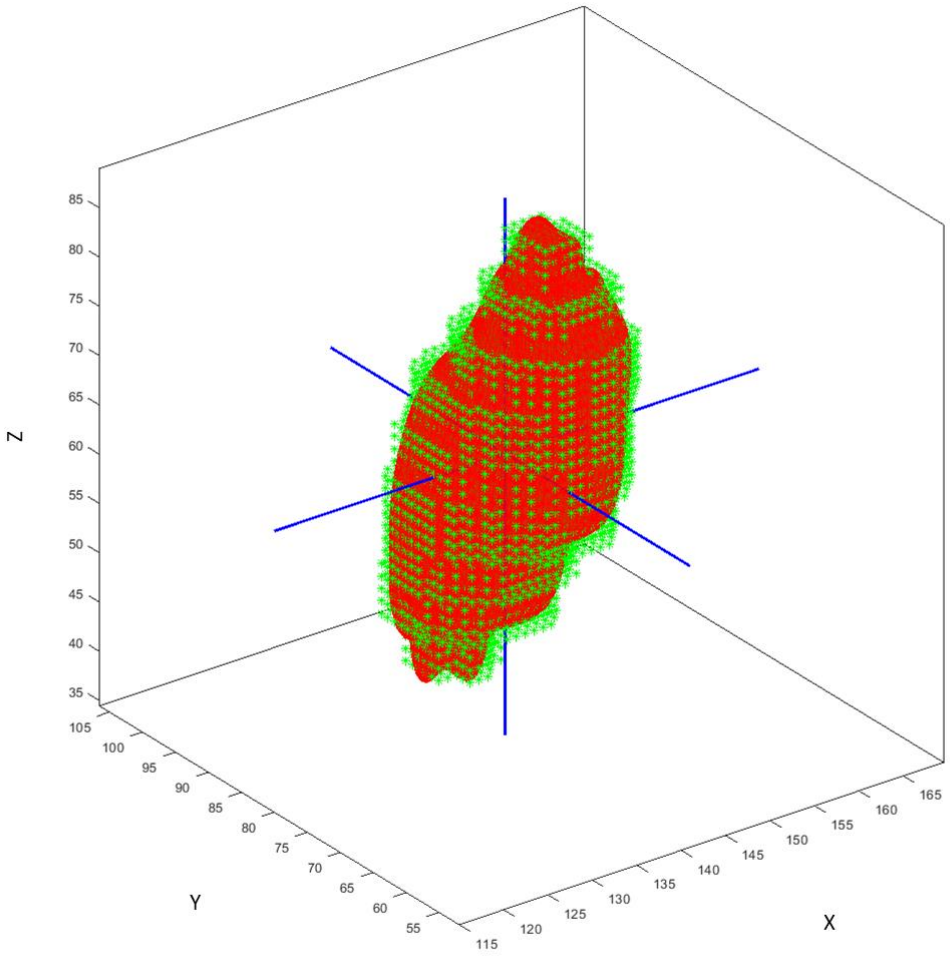

Figure 5.5. Overlap image of the chromosome sample (in green) and its reconstructed surface (in red) shown in Figure 5.4.

To access to the performance of the SPHARM surface modeling, the root mean square error described in section 3.4.2 is computed for all the sample objects including chromosomes and nuclei. Table 5.1 shows the average, standard deviation, sample size and the 95 % confidence interval of chromosomes and nuclei. With a total number of 27 normal diploid cells and 34 aneuploid cells (trisomy 12), there are 122 sample objects of

chromosome 8 and X, 156 sample objects of chromosome 12, and 61 nuclei sample objects. The unit of RMSE is in per pixel and the pixel size of our image is 0.059 um/pixel.

Table 5.1. Average and statistic assessment of the RMSE between input objects and reconstructed surfaces.

	Chr-8	Chr-X	Chr-12	Nucleus
# of samples	112	112	141	56
Average RMSE	1.68	1.53	1.66	3.84
STD	0.98	0.61	0.94	1.64
95%CI	±0.17	±0.10	±0.14	±0.41

The average RMSE of chromosomes are between 1.53 to 1.68 pixels, which is around 0.090 to 0.099 um. The nuclei samples have an average RMSE of 3.84 that converted to 0.22 um. As shown in Figure 5.6, the boxplot shows the 50, 75, and 25 percentiles at the middle, top, and bottom line of each box respectively. The whiskers extend to the maximum and minimum values excluding the outliers. The red "+" indicates the outliers. The RMSE distributed similarly for chromosome 8, X, and 12. The RMSE values for nuclei are higher than chromosomes that may due to the size of a nucleus is much larger than the chromosomes within it. The number of outliers of chromosome 8, X, 12, and nuclei are 9, 8, 7, and 4 respectively.

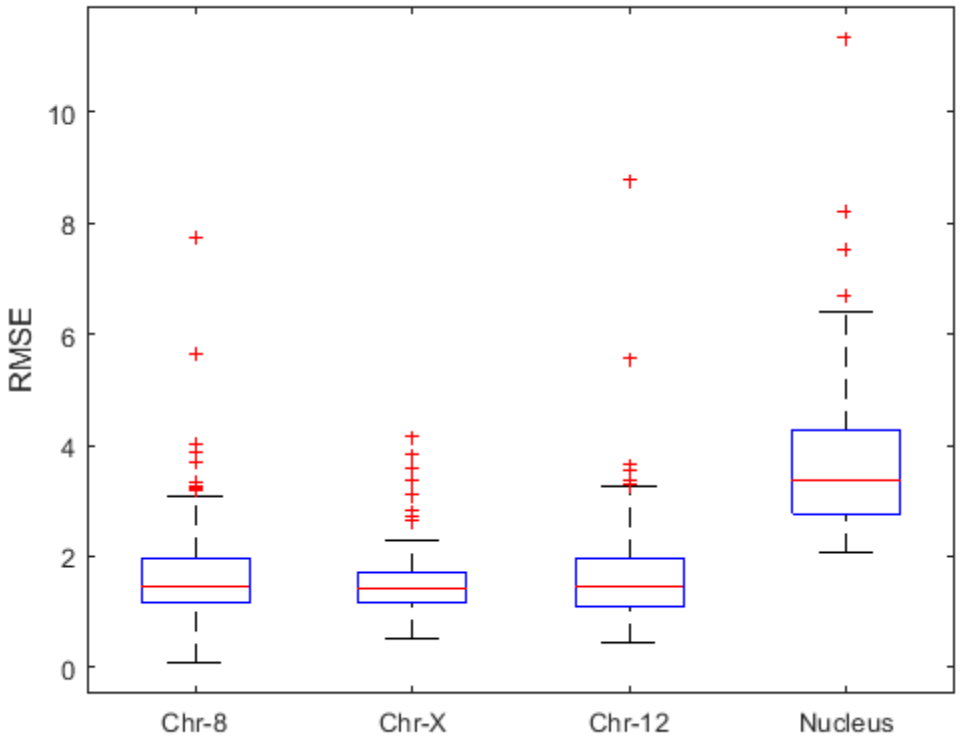

Figure 5.6. Root square mean error between input object and reconstructed surface of chromosomes and nuclei.

After reconstructing the surface of all sample objects (i.e., nuclei and chromosomes within them), the objects that belongs to the same sample are reassembled. Figure 5.7 gives example of reconstructed nuclei and the chromosomes within them and the corresponding FISH image. Figure 5.7 (a) and (b) are maximum intensity projection of optical sections of a diploid nucleus and a trisomy 12. The nuclei are colored in blue and chromosome 8, X, and 12 are colored in yellow, green, and red respectively. Figure 5.7 (c) and (d) are 3D mesh surfaces of chromosomes and nuclei that are reconstructed using SPHARM shape descriptor with the same color code as Figure 5.7 (a) and (b).

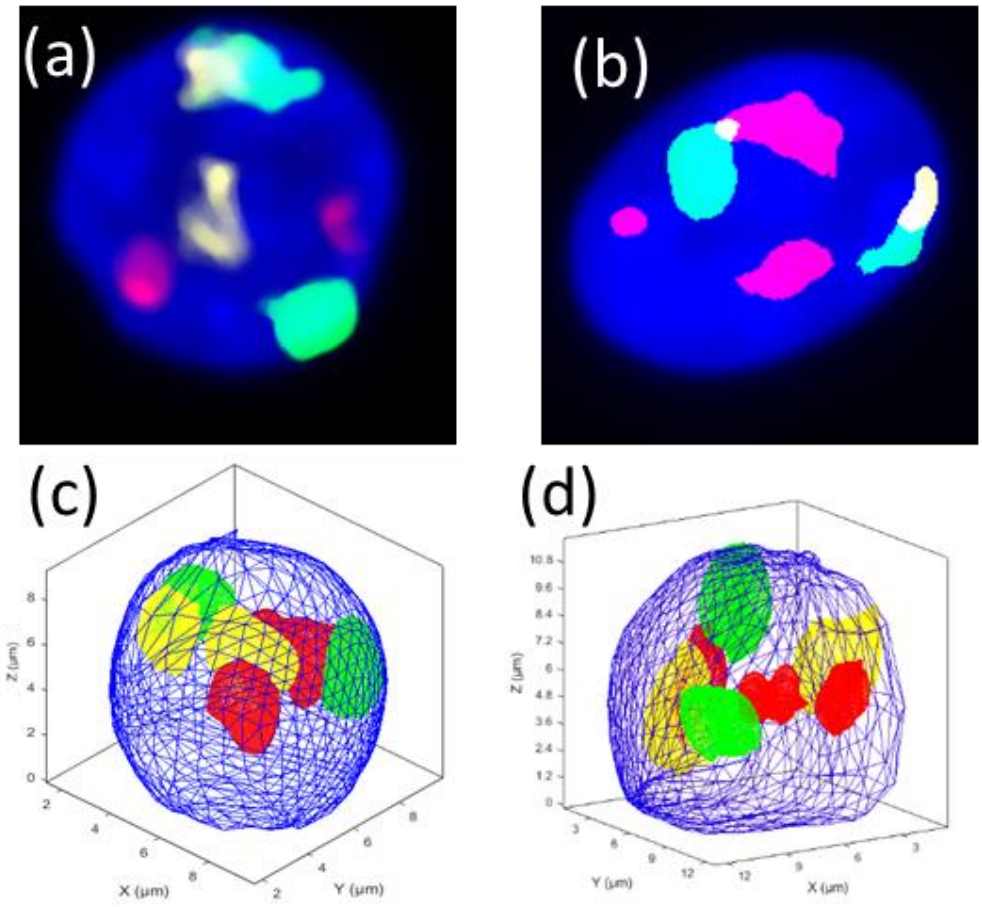

Figure 5.7. The maximum intensity projection of optical sections and the SPHARM reconstructed surface.

5.2 CTs Spatial Organization

One critical factor that needs to be considered while defining 3D position of any object is to establish a reference frame or structure relative to the space that is specified. In this study, the orientation of each nucleus can vary depending on how the cells are fixed on the sample slide and thus we have to align all nuclei samples with their enclosed chromosomes to a common coordinate framework. The alignment is based on the shape of nuclei and the shape information is extracted by spherical harmonic decomposition. The SPHARM coefficients contain information on the shape of the modeled object and we align each nucleus to a template by minimizing the differences of SPHARM coefficients between the

nuclei and the template. The template SPHARM model was generated by finding a shape that had the minimum sum of RMSD to corresponding SPHARM coefficients of all nuclei.

The 3D modeling framework proposed in this study serves as a tool for quantifying the spatial localization of chromosome territories within the nucleus. It allows us to retrieve the relative positioning, size, and shape information of chromosome territories and nuclei. The following analysis is performed on each sample nucleus in order to quantify the spatial organization of chromosome territories, and study the CT proximity relationships in 3D space. The following measurements are performed: 3D positioning of CTs, radial distance, intra-homologous and inter-heterologous chromosome distances, and CT volume distribution.

5.2.1 Three Dimensional positioning of CTs

The relative 3D position of each homologous chromosome with respect to the local landmark is retrieved as mentioned in the previous chapter. Local landmark in nucleus is determined by the proximity with respect to the selected chromosome in the reference nucleus sample (global landmark). Visualizing the spatial position of each chromosome gives a direct insight into CT organization within nucleus. The position of CTs are presented in a (x, y, z) coordinate system with the origin point at the center of mass of the nucleus.

Chromosome X

The chromosome X, homolog group 1 in each cell sample is selected as the local landmark that is used as the reference to determine the position of other chromosomes. To evaluate the 3D position of chromosome X-2 relative to chromosome X-1, we perform the

student's t-test to compare the 3D position of CTs with respect to the local landmark. The position of CTs in (x, y, z) coordinate is transformed to a single value by computing the scalar projection of the CTs relative to the local landmark for the statistical analysis. The scalar projection of a vector $\boldsymbol{b}$ on vector $\boldsymbol{a}$ (defined in section 3.6.3) contains both magnitude and directional information. It gives the length of the orthogonal projection of the vector $\boldsymbol{b}$ onto the vector $\boldsymbol{a}$. If the two vectors are along the same direction the scalar value is positive and it is negative when they are at opposite directions. The student's t-test is performed under the null hypothesis that the mean of two groups (diploid and aneupolid nucleus) are the same with equal variances using a 95% confidence interval.

Table 5.2. Student's t-test of mean position for chromosome X-2.

Chromosome X-2						
	Mean	Variance	Observations	t Stat	P(T<=t)	t Critical
Aneuploid	0.026776	0.020074	29	1.366192	0.177541	2.004879
Diploid	-0.01831	0.010003	27			

As shown in Table 5.2, the p-value for comparing normal diploid and aneuploid cells is greater than 0.05 which indicates that the test fails to reject the null hypothesis for equal mean of the two groups.

The 3D positioning of the centroid (mass center) of chromosome X, homolog 2 respect to the local land mark is shown in Figure 5.8. The black cross symbol represents the local landmark (chromosome X-1) in each cell sample that has been aligned and scaled to the global landmark. Therefore, it only shows at one position for all diploid and aneupolid samples. The green dots are the centroid position of chromosome X-2 respect to the reference chromosome X-1.

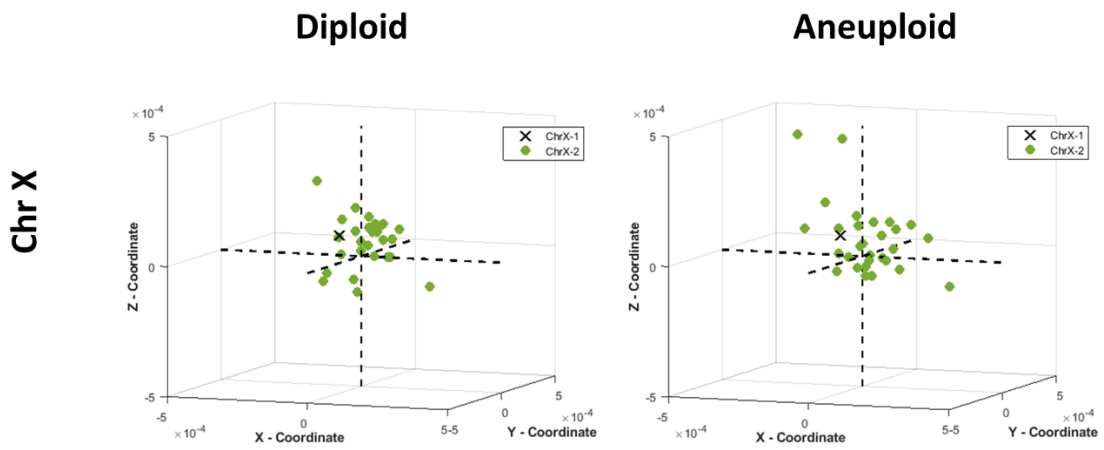

Figure 5.8. Spatial position in 3D space of the centroid of chromosome X-2 (diploid and aneuploid cells) relative to the landmark chromosome X-1.

As seen in Figure 5.8, the spatial distribution of chromosome X-2 are not significantly different between diploid and aneupolid samples (p-value > 0.05). To estimate and visualize the CT clustering across samples, the 3D distribution of chromosome X-2 centroids is represented using ellipsoid as shown in Figure 5.9. The position of the ellipsoid is determined by the mean location of chromosome X-2 and the reference coordinate (orientation of the ellipsoid) is specified with the principle components of the position of chromosome X-2. The size of ellipsoid along each principle components axis is defined by the standard deviation of each principle component. As seen in both Figure 5.8 and Figure 5.9, the chromosome territory for chromosome X-2 relative to the landmark chromosome X-1 are similar for both diploid and aneupolid cell samples, which indicates the non-random chromosome territories distribution for chromosome X-2 is not affected by trisomy 12. The CTs for chromosome X-2 in diploid and aneupolid cells occupy similar space within the nucleus.

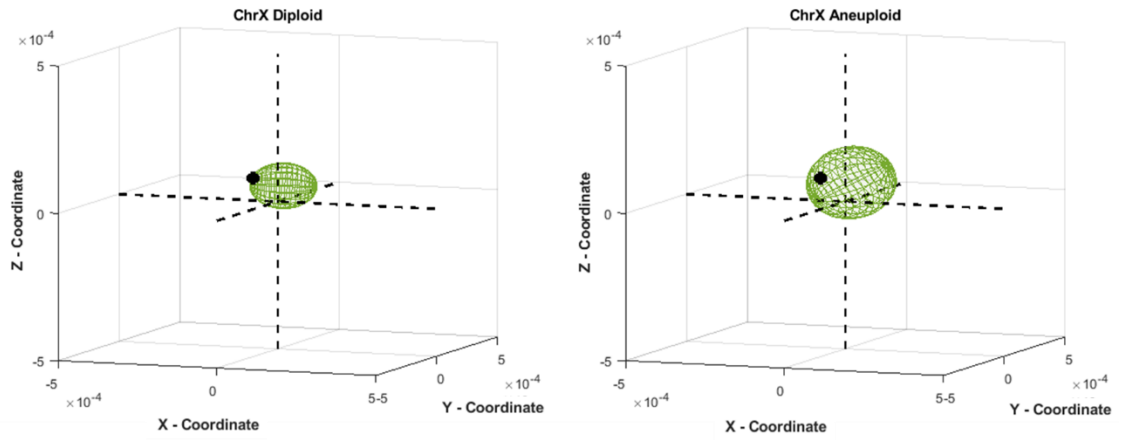

Figure 5.9. Visualization of the clustering of chromosome X-2's 3D position respect to the reference chromosome for both diploid and aneuploid samples.

In order to give a better view of the clustering distribution of chromosome X-2 relative to reference chromosome X-1, we reduce the dimension using principle component analysis and find the 2D Gaussian of the first and second principle components shown in Figure 5.10, which also shows a similar distribution between normal diploid cells and aneupolid cells.

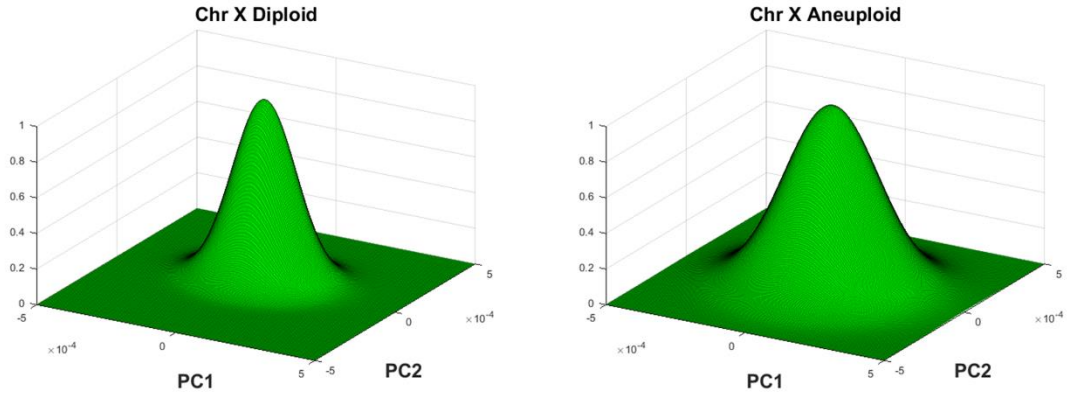

Figure 5.10. Two dimensional Gaussian estimation of clustering distribution of chromosome X-2 relative to the landmark chromosome X-1 for diploid and aneuploid nuclei.

Based on the visualization and statistical assessment of our data described above, the position of chromosome X, homolog 2 occupies a similar space within the nucleus for both normal and aneuploid cells. This indicates that the chromosome territories are distributed

in a non-random fashion and each of the CTs is located at an unique position as suggested in previous publications [5], [63]. Also we noted that the similar positioning of chromosome X-2 for normal and aneuploid nucleus shows that even the cell samples gain an extra chromosome 12 (trisomy 12), it does not change the chromosome territory of chromosome X.

Chromosome 8

Here we compare the mean of scalar projection value of chromosome 8-1 and chromosome 8-2 with respect to the landmark chromosome X-1 in normal and aneuploid samples. We performed the following tests for assessing statistical significance, (1) test the difference in mean projection value within the same homolog group (i.e., aneuploid vs diploid in chromosome 8-1 and chromosome 8-2) and (2) test the difference between the homolog groups (i.e., chromosome 8-1 vs chromosome 8-2 in aneuploid and diploid samples). The results of student's t-test of chromosome 8 with equal variance and 95% confidence interval are shown in Table 5.3.

Table 5.3. Student's t-test of mean position for chromosome 8-1 and chromosome 8-2 in diploid and aneuploid nucleus.

Aneuploid vs Diploid (Chromosome 8-1)						
	Mean	Variance	Observations	t Stat	P(T<=t)	t Critical
Aneuploid	0.083271	0.010048	29	0.174458	0.862158	2.004879
Diploid	0.079242	0.004663	27			
Aneuploid vs Diploid (Chromosome 8-2)						
Aneuploid	-0.09544	0.010027	29	-1.22557	0.225677	2.004879
Diploid	-0.06501	0.007097	27			
Chromosome 8-1 vs Chromosome 8-2 (Aneuploid)						
Chromosome 8-1	0.083271	0.010048	29	6.792154	7.63E-09	2.003241
Chromosome 8-2	-0.09544	0.010027	29			
Chromosome 8-1 vs Chromosome 8-2 (Diploid)						
Chromosome 8-1	0.079242	0.004663	27	6.911892	6.84E-09	2.006647
Chromosome 8-2	-0.06501	0.007097	27			

The p-values for within homolog group in chromosome 8-1 and chromosome 8-2 are 0.44 and 0.64 respectively, which both indicate a failure (p-value > 0.05) for rejecting the null hypothesis (equal mean). This result suggests that CTs occupy similar position within nucleus. This finding is consistent with our visual observation from the distribution plots above and also show the non-random spatial distribution of CTs. The "between groups" test in aneuploid and diploid have p-values < 0.05. For both these tests the null hypothesis is rejected indicating the significance of differences in mean position. This result confirms the unique positioning of CTs for each homolog of chromosome 8 (i.e., each chromosome occupies a unique space within nucleus).

The spatial distribution of chromosome 8 in both homolog 1 and 2 can be visualized and analyzed with the same method that we have mentioned in the previous section. Figure 5.11 shows the mass center distribution for chromosome territories of chromosome 8, homolog 1 and 2, in normal and aneupolid nucleus. In both cases (normal diploid and

aneuploid cells), the centroid within the same homologous group tend to be clustered together and the spatial distribution also shows a high similarity between normal cells and trisomy 12.

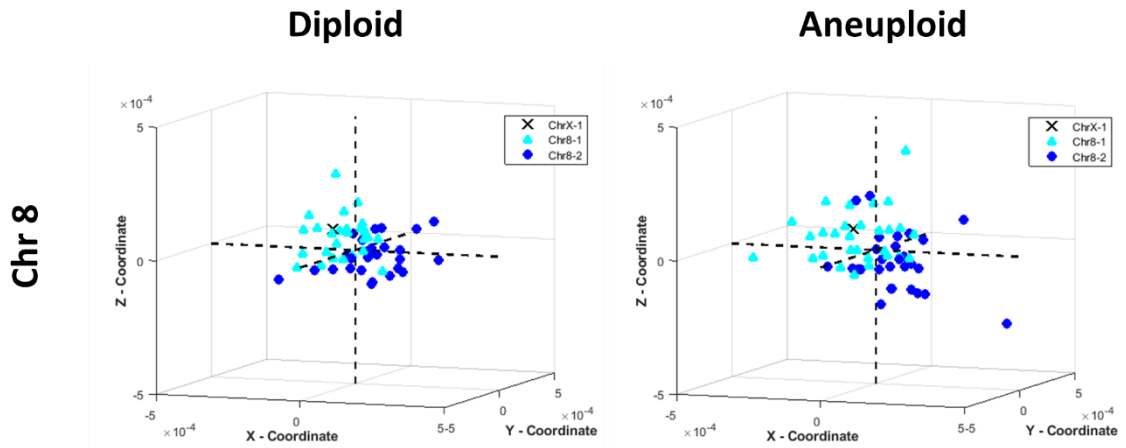

Figure 5.11. Spatial position in 3D space of the centroid of chromosome 8-1 and 8-2 (diploid and aneuploid cells) relative to the landmark chromosome X-1.

The 3D ellipsoids in Figure 5.11 provides a visual representation of the spatial clustering of chromosome 8-1 and 8-2 with respect to the local landmark (chromosome X-1) for both normal and aneuploidy samples. The result shows a distinct spatial position in 3D space within the nucleus between chromosome 8-1 and 8-2 in both diploid and aneuploidy cells. It also illustrates to the idea of non-random spatial organization of CTs by showing a similar position in chromosomes for the same homologous group in diploid and aneuploid samples.

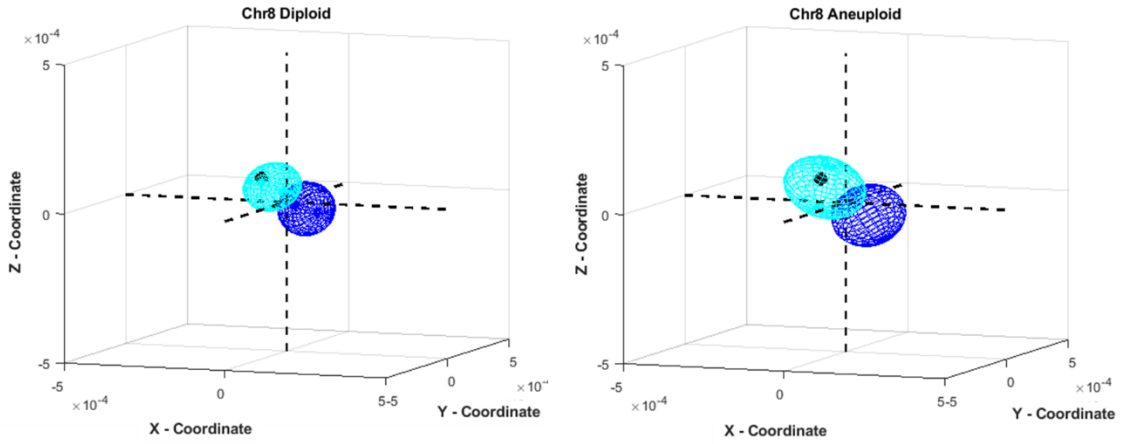

Figure 5.12. Visualization of the clustering of chromosome 8-1 and 8-2's 3D position respect to the reference chromosome for both diploid and aneuploid samples.

The 2D Gaussian distribution of the positioning of chromosome 8 is generated with the PCA to show the most and the second most variability axes, as seen in Figure 5.13. The red curve indicates the 2D Gaussian distribution of chromosome 8 in homolog group 1 and green curve represents the homolog group 2 of chromosome 8. The distribution of the first and second principle components suggest a unique positioning between homologous group 1 and 2 in chromosome 8 along with similar localization in the same homologous group in normal and abnormal cell samples.

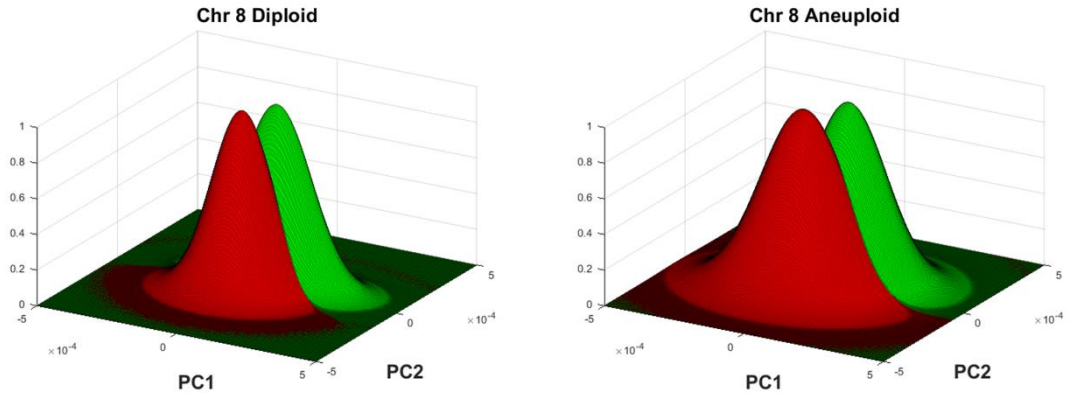

Figure 5.13. Two dimensional Gaussian estimation of clustering distribution of 8-1(red) and 8-2(green) relative to the landmark chromosome X-1 for diploid and aneuploid nuclei.

Chromosome 12

Since trisomy 12 has an extra chromosome 12, here we conduct the student's t-test on all pairings of the different chromosome 12 homologs to test the significance of difference between aneuploid and diploid cell samples. Table 5.4 shows all the statistic comparison of chromosome 12 in both diploid and aneuploid samples. Student's t-test with a 95% confident interval is used for comparing each individual chromosome 12 between normal sample group and aneuploid sample group. It also used to test the significance level of different means between chromosome 12-1 and 12-2 within the diploid sample group. To compare the position of all the three homolog of chromosome 12 in trisomy 12 cells, analysis of variance (ANOVA) is used. ANOVA is a statistical test that provides a test to compare whether the means of several groups are equal or not. One interesting finding is that only two groups out of all the comparisons have a p-value > 0.05 which shows no significant difference in the position, they are:

(1) aneuploid(chromosome 12-1) vs diploid (chromosome 12-1) and

(2) aneuploid(chromosome 12-3) vs diploid (Chromosome 12-2).

As we compare the position of chromosome 12-1 in normal and abnormal sample group (p-vale = 0.04) we the null hypothesis indicating similar position of the homolog across the two samples. Despite, chromosome 12-3 in aneuploid sample and chromosome12-2 in normal samples are not labeled as the same homolog group, their CT positions are not significantly different. Since we do not know which homolog of the chromosome 12 is the extra one in the trisomy 12 and the labeling is done relative to the proximity of the chromosomes to the local landmark, the chromosome 12-3 in aneuploid samples and

chromosome 12-2 in diploid samples may actually belongs to the same homologous group. The position difference between chromosome 12-2 from aneuploid group and chromosome 12-1 from the normal group is not significant (p-value >0.5). This indicates that these two groups belong to the same homolog of the chromosome 12. All other pairings of chromosome 12 (including ANOVA) show significant difference in positioning (p-value < 0.05) which suggest the uniqueness of the position of each chromosome within the nucleus.

Table 5.4. Student's t-test and ANOV test of mean position for chromosome 12-1, chromosome 12-2 in diploid and extra chromosome 12-3 in aneuploid nucleus.

	Mean	Variance	Observations	t/f Stat	P(T<=t)	t/f Critical
Aneuploid(Chromosome 12-1) vs Diploid (Chromosome 12-1) using t-test						
Aneuploid(12-1)	0.117898	0.006548	29	2.102833	0.040157	2.004879
Diploid(12-1)	0.073981	0.005615	27			
Aneuploid(Chromosome 12-1) vs Diploid (Chromosome 12-2) using t-test						
Aneuploid(12-1)	0.117898	0.006548	29	8.299342	3.25E-11	2.004879
Diploid(12-2)	-0.05065	0.004925	27			
Aneuploid(Chromosome 12-2) vs Diploid (Chromosome 12-1) using t-test						
Aneuploid(12-2)	0.030535	0.010313	29	-1.81056	0.075774	2.004879
Diploid(12-1)	0.073981	0.005615	27			
Aneuploid(Chromosome 12-2) vs Diploid (Chromosome 12-2) using t-test						
Aneuploid(12-2)	0.030535	0.010313	29	3.455413	0.001078	2.004879
Diploid(12-2)	-0.05065	0.004925	27			
Aneuploid(Chromosome 12-3) vs Diploid (Chromosome 12-1) using t-test						
Aneuploid(12-3)	-0.0562	0.007275	29	-6.04924	1.42E-07	2.004879
Diploid(12-1)	0.073981	0.005615	27			
Aneuploid(Chromosome 12-3) vs Diploid (Chromosome 12-2) using t-test						
Aneuploid(12-3)	-0.0562	0.007275	29	-0.26484	0.792141	2.004879
Diploid(12-2)	-0.05065	0.004925	27			
Chromosome 12-1 vs chromosome 12-2 (Diploid group) using t-test						
Chr 12-1	0.073981	0.005615	27	6.308005	6.26E-08	2.006647
Chr 12-2	-0.05065	0.004925	27			
Chromosome 12-1 vs chromosome 12-2 (Aneuploid group) using t-test						
Chr 12-1	0.117898	0.006548	29	3.623134	0.000628	2.003241
Chr 12-2	0.030535	0.010313	29			
Chromosome 12-1 vs chromosome 12-3 (Aneuploid group) using t-test						
Chr 12-1	0.117898	0.006548	29	7.974273	8.52E-11	2.003241
Chr 12-3	-0.0562	0.007275	29			
Chromosome 12-2 vs chromosome 12-3 (Aneuploid group) using t-test						
Chr 12-2	0.030535	0.010313	29	3.522128	0.00086	2.003241
Chr 12-3	-0.0562	0.007275	29			
All three Chromosome 12 (Aneuploid group) using ANOVA						
Chr 12-1	0.117898	0.006548	29	27.31473	7.27E-10	3.105157
Chr 12-2	0.03053	0.01031	29			
Chr 12-3	-0.0562	0.00727	29			

About half of our samples are normal diploid cells and the other half samples are aneuploid which have gained an extra chromosome 12 (trisomy 12). Figure 5.14 plots out the mass center of chromosome 12-1, 12-2, and 12-3 (for aneuploid). As seen in the figure, the chromosome in the same homolog group tens to cluster together in both diploid and aneuploid cell samples.

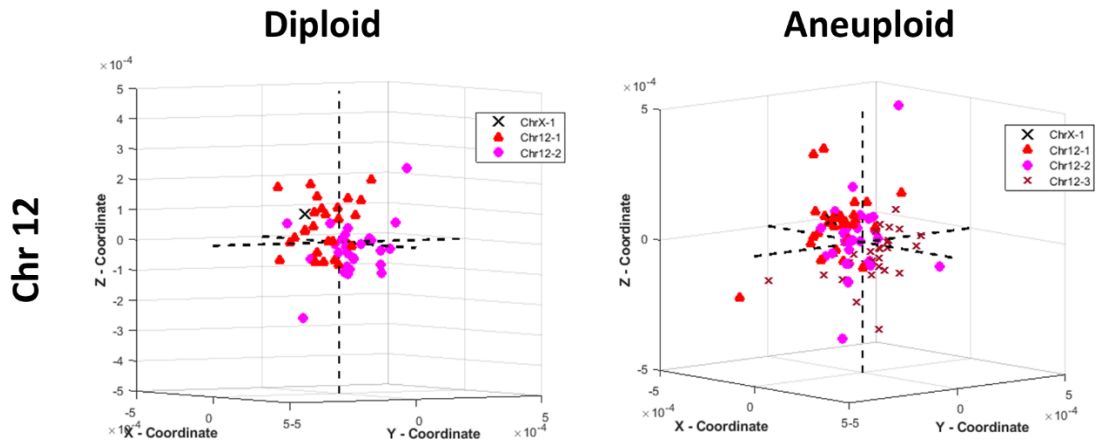

Figure 5.14. Spatial position in 3D space of the centroid of chromosome 12-1, 12-2, and 12-3 (aneuploid cells) relative to the landmark chromosome X-1.

The estimated CTs position ellipsoid of chromosome 12 relative to the landmark, chromosome X-1, is shown in Figure 5.15 and the 2D Gaussian distribution of CT position after PCA dimension reduction is shown in Figure 5.16. We can observe the distinct territories for each individual chromosome in both normal and abnormal samples form the CT ellipsoid visualization and the 3D Gaussian distribution estimation. However it is hard to visually spot if there is any similarity in terms of CT positioning between diploid and aneuploid samples in the scatter plots.

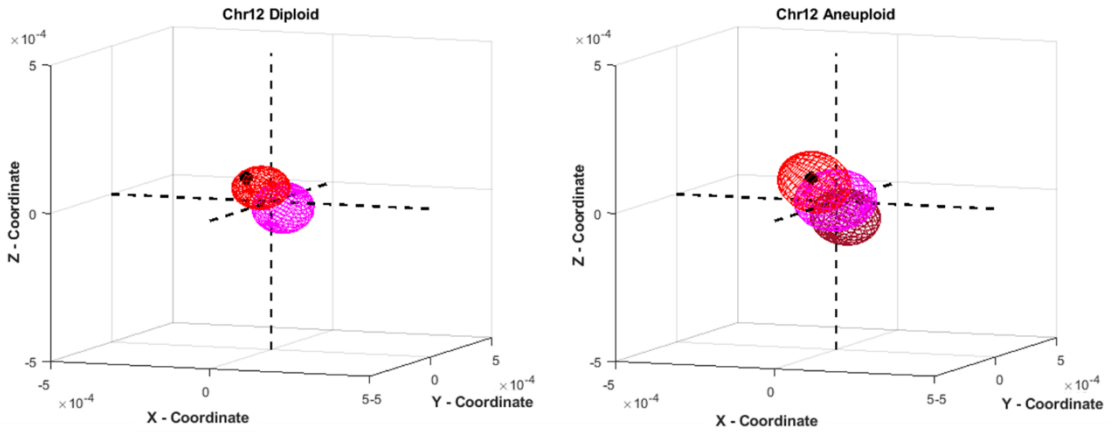

Figure 5.15. Visualization of the clustering of chromosome 12-1, 12-2, and 12-3's (aneuploid) 3D position respect to the reference chromosome for both diploid and aneuploid samples.

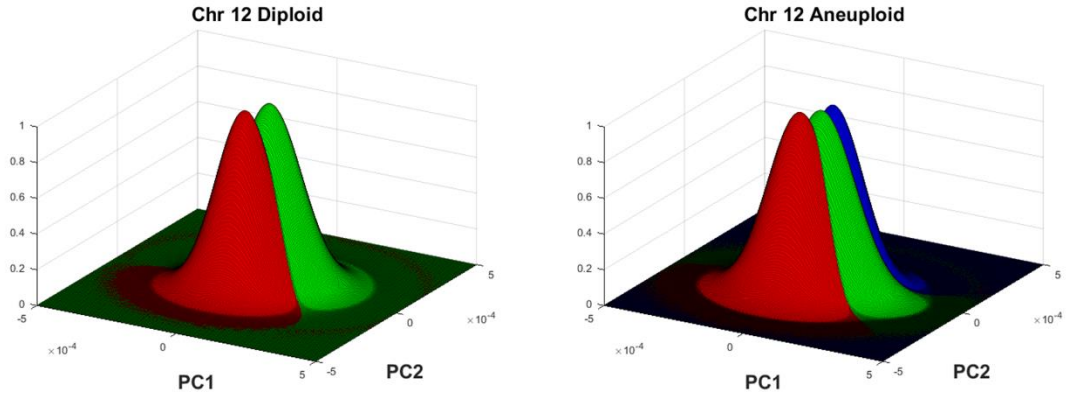

Figure 5.16. Two dimensional Gaussian estimation of clustering distribution of 12-1(red), 12-2(green), and 12-3(blue for aneuploid) relative to the landmark chromosome X-1 for diploid and aneuploid nuclei.

Overall CT localization comparison

Figure 5.17 shows the average of the centroid of each chromosome relative to the local landmark, chromosome X-1 (the black cross in the figure), in normal diploid sample group and aneuploid sample group. The same color code indicates the chromosomes that are labeled as the same homolog group. The plus shape markers belong to aneuploid sample group and the dots represent the diploid group. The markers that are circled together indicate the instances where we fail to reject the null hypotheses. As seen in the figure, the

normal and abnormal group of chromosome X-2, 8-1, and 8-2 are located at similar position. Chromosome 12-2 in normal group and chromosome 12-3 in abnormal group are at similar position within nucleus. Also chromosome 12-1 in normal group and chromosome 12-2 in aneuploid group share similar position. The position of chromosome12-1 in the aneuploid sample group is significantly different from all other chromosome.

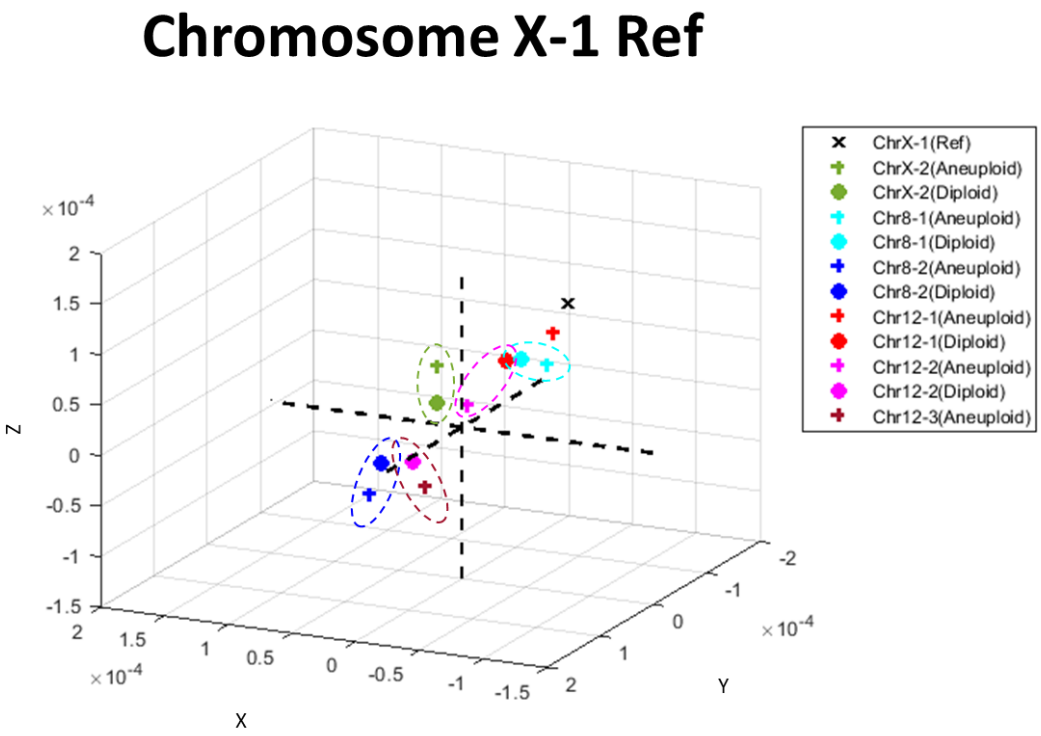

Figure 5.17. Average 3D position of mean centroid of each individual chromosome in normal diploid cells and aneuploid cells.

The 3D position of homologs was analyzed by pairwise comparison within diploid and aneuploid group. The test was conducted using paired t-test with zero mean null hypothesis and a confidence level of 95%. The result p-values are listed in Table 5.5. Tests where we fail to reject the null hypothesis are labeled with "#" sign which indicates the homologs have a close proximity.

Table 5.5. Paired t-test of 3D position of CTs in diploid and aneuploid cell populations.

	Aneuploid						
	ChrX-1	ChrX-2	Chr8-1	Chr8-2	Chr12-1	Chr12-2	Chr12-3
ChrX-1	1.00E+00						
ChrX-2	5.79E-05	1.00E+00					
Chr8-1	1.06E-03	5.53E-02#	1.00E+00				
Chr8-2	1.35E-13	2.77E-03	5.96E-07	1.00E+00			
Chr12-1	3.48E-02	2.07E-04	9.71E-02#	2.24E-08	1.00E+00		
Chr12-2	6.26E-07	8.97E-01#	8.17E-02#	5.00E-05	1.01E-05	1.00E+00	
Chr12-3	1.84E-13	8.54E-03	1.02E-05	9.10E-02#	8.66E-11	1.62E-08	1.00E+00
	Diploid						
	ChrX-1	ChrX-2	Chr8-1	Chr8-2	Chr12-1	Chr12-2	
ChrX-1	1.00E+00						
ChrX-2	2.77E-09	1.00E+00					
Chr8-1	9.55E-06	2.39E-04	1.00E+00				
Chr8-2	3.89E-13	3.97E-02	3.78E-08	1.00E+00			
Chr12-1	1.31E-05	1.74E-03	7.62E-01#	2.02E-05	1.00E+00		
Chr12-2	2.79E-14	1.96E-01#	2.52E-07	5.12E-01#	1.42E-08	1.00E+00	

To ensure the proposed alignment method and the choice of template (average shape) is reliable, we compared the position analysis using average shape as template and one of the nucleus as template. Table 5.6 gives the p-value of t-test and ANOVA on the CT position with average shape and nucleus sample as template. For chromosome X and 8, the two template showed a similar result that homologs of chromosome 8 and X occupied specific space within nucleus and does not affected by aneuploidy. For chromosome 12, in both templates, there were two homologs located at similar position between diploid and aneuploid samples and one homolog that are away from the other two.

Table 5.6. The p-value of student's t-test on the projection values of chromosome8, X, and 12 between normal and abnormal sample groups using average shape and nucleus sample as template.

	Chromosome X			
Abnormal:Normal	X-1:X-1	X-1:X-2	X-2:X-1	X-2:X-2
Average shape	1#	1.51834E-12	2.95795E-05	0.17754134#
Nucleus sample	1#	1.03776E-11	9.98878E-09	0.328092214#
	Chromosome 8			
Abnormal:Normal	8-1:8-1	8-1:8-2	8-2:8-1	8-2:8-2
Average shape	0.862157896#	1.90867E-07	4.88977E-10	0.225677246#
Nucleus sample	0.376624984#	3.68547E-08	1.62752E-08	0.851742171#

	Average shape			**Nucleus sample**		
	Chr12-1 Abnormal	Chr12-2 Abnormal	Chr12-3 Abnormal	Chr12-1 Abnormal	Chr12-2 Abnormal	Chr12-3 Abnormal
Chr12-1 Normal	4.01E-02	7.57E-02#	1.42E-07	9.90E-02#	1.26E-02	4.99E-08
Chr12-2 Normal	3.22E-11	1.07E-03	7.92E-01#	5.02E-10	4.44E-03	6.61E-01#
	Average shape			**Nucleus sample**		
Chr 12-1 vs 12-2 (Diploid, t-test)	6.26E-08			1.69E-07		
(Aneuploid, ANOVA)	7.27E-10			5.05E-10		

5.2.2 Radial and peripheral distance

The radial distance is obtained as described in chapter 3. Figure 5.18 is the bar plot of the average radial distance of each chromosome in normal and abnormal sample group. For chromosome X and 12, the radial distance show similar trend in both normal and abnormal sample group. Student's t-test also shows no significance in difference between normal and abnormal samples for the same homolog group. However, the radial distance of chromosome 8-2 in trisomy 12 is statistically greater than chromosome 8-2 in normal samples.

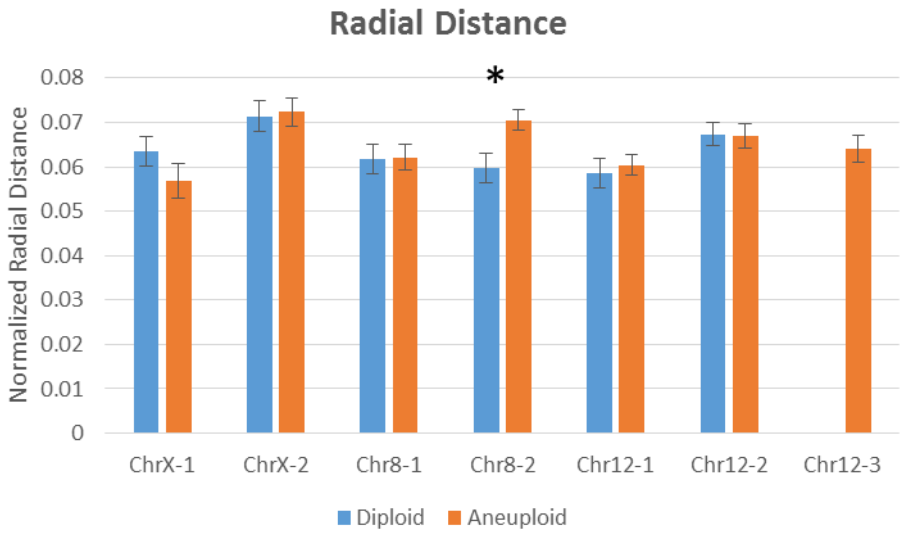

Figure 5.18. Average radial distance of each homolog in chromosome X, 8, and 12. The asterisk symbol indicates the two-sample student's t-test shows significantly different between the two groups at 95% confidence.

Since the labeling may be different for chromosome 12 in diploid and aneuploid group, we compared the mean radial distance with all possible pairs of chromosome 12 between the diploid and aneuploid. As the t-test result of all comparison listed in Table 5.7, the radial distance show no significant difference between diploid and aneuploid group.

Table 5.7. The p-value of t-test on radial distance of homologs of chromosome 12 in diploid and aneuploid group.

Chromosome	**12-1 (aneuploid)**	**12-2 (aneuploid)**	**12-3 (aneuploid)**
12-1 (diploid)	0.647587726	0.055925484	0.222097684
12-2 (diploid)	0.051909289	0.944269397	0.419743673

We investigated the radial distance between the homologs that belong to the same chromosome and see if there are any alterations between diploid and aneuploid group. Table 5.8 lists the p-value result from student t-test of all the pairing within diploid and aneuploid group. The "#" indicates where the p-value show significance (p-value<0.05) and the ones that were significantly different in radial distance and belong to the same

chromosome were highlighted. In diploid group, there was no significant difference between the two homologs in chromosome X and 8. However, significant difference in the mean radial distance of homologs in chromosome X and 8 was observed in aneuploid group. Moreover, homologs in chromosome 12 show significant difference in terms of radial distance in diploid group but not in the aneuploid group.

Table 5.8. Student t-test on the mean radial distance of homologs of chromosome X, 8, and 12 within diploid and aneuploid group.

Radial Distance P-value in Aneuploid							
Chr	X-1	X-2	8-1	8-2	12-1	12-2	12-3
X-1	NA	0.002624#	0.277254	0.00318	0.433915	0.036058#	0.14285
X-2		NA	0.020545#	0.001984#	0.003079#	0.209789	0.059212
8-1			NA	0.027899#	0.638026	0.234871	0.650643
8-2				NA	0.002688#	0.337068	0.089323
12-1					NA	0.072528	0.333741
12-2						NA	0.470286
12-3							NA
Radial Distance P-value in Diploid							
Chr	X-1	X-2	8-1	8-2	12-1	12-2	
X-1	NA	0.105763	0.719619	0.433061	0.304045	0.373847	
X-2		NA	0.051789	0.018476#	0.009988#	0.346123	
8-1			NA	0.674871	0.506663	0.202305	
8-2				NA	0.803216	0.080048	
12-1					NA	0.044420#	
12-2						NA	

5.2.3 Intra-homologous and inter-heterologous distances

To investigate the relationship across different chromosomes as well as homolog of the same chromosome, we computed inter-heterologous (distance between different chromosomes) and intra-homologous (distance between homolog in the same chromosomes) distances as mentioned in chapter 3. Figure 5.19 presents the inter-heterologous distance of chromosome X to chromosome 8, chromosome 8 to chromosome

12, and chromosome 12 to chromosome X for diploid and aneuploid sample groups. The asterisk mark indicates where the stutent's t-test show significant differences between normal and abnormal samples. The inter-heterologous distance of chromosome 12 to chromosome 8 and chromosome X have significantly decreased in abnormal sample group comparing to normal sample group. These changes not only happen when the aneuploid (chromosome 12) is involved, but also appear in the distance between other chromosomes. The inter-heterologous distance between chromosome X and chromosome 8 shows a significant decrease in aneuploid group. The differences in inter-heterologous distance observed in the trisomy 12 cells may have been influenced by the changes in CT positions due to the extra chromosome 12. These changes not only affect the aneuploid chromosome (chromosome 12), but could also influence other chromosomes (in this case, chromosome 8 and/or chromosome X).

Intra-homologous distance measures the distance between the homologs of one chromosome. Since there are three combinations of intra-homologous distances for trisomy 12, the intra-homologous distance for trisomy 12 is defined by the minimum of the three combinations. The intra-homologous distance shows no significance in mean distance between normal and abnormal sample groups in chromosome 8 and chromosome X. On the other hand, the homologous distance of chromosome 12 is significantly decreased when gaining one extra chromosome 12 (as indicated in Figure 5.20 with asterisk).

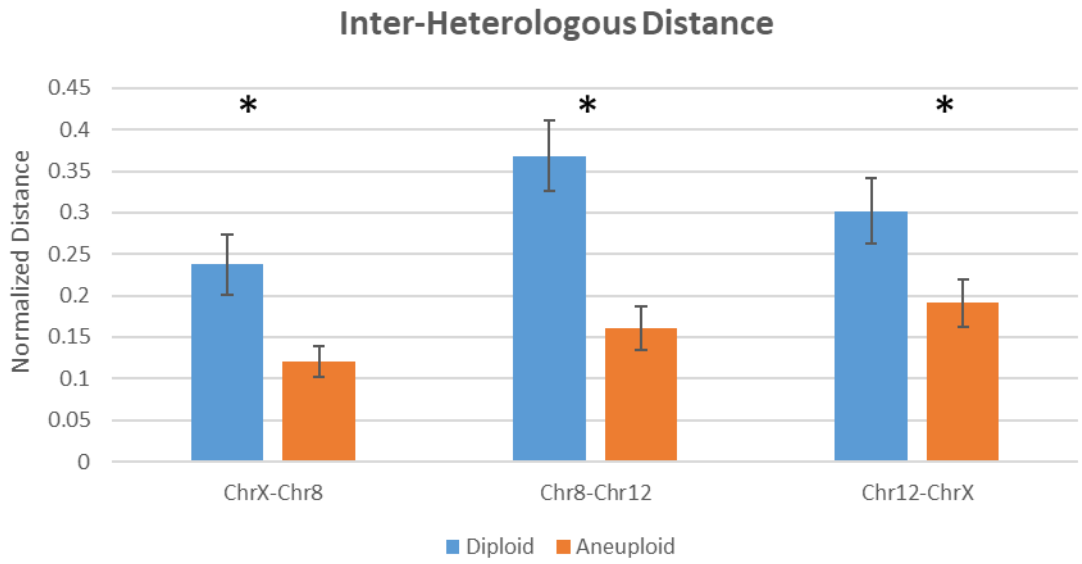

Figure 5.19. Inter-heterologous distance of chromosome X, 8, and 12 in normal diploid and aneuploid sample groups.

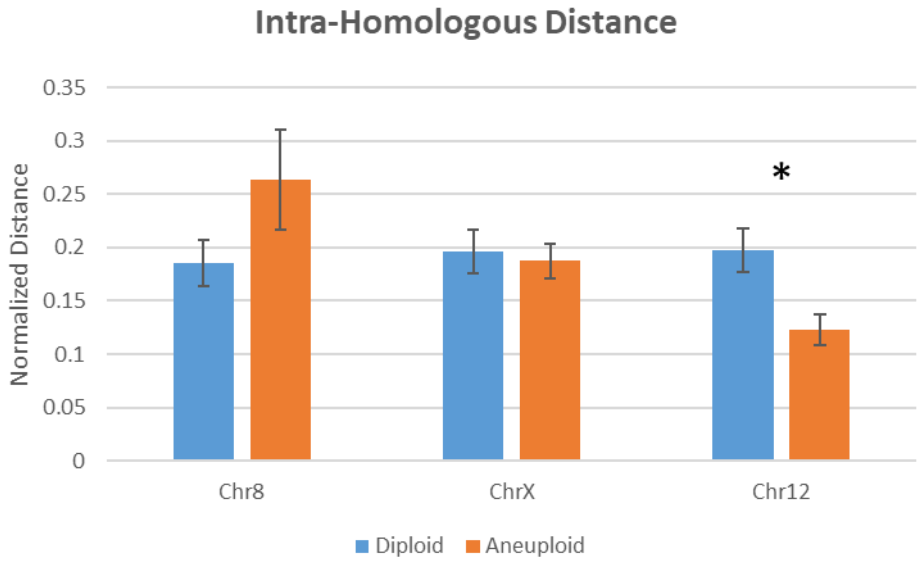

Figure 5.20. Intra-homologous distance of chromosome X, 8, and 12 in normal diploid and aneuploid sample groups.

5.2.4 CT Volume

The CT volume is computed as mention in chapter 3 for each chromosome territory. The CT volume of each sample is normalized respect to the volume of the nucleus of each sample. Figure 5.21 shows the mean value of CT volume for each homolog chromosome in normal group and abnormal sample group. Two sample student's t-test assuming equal variance with 95% confidence level is used to test the changes in CT volume between normal and abnormal group. Chromosome 12 and chromosome 8 show the tendency of decreasing in volume for all homologs, but the change is not recognized as significant. The volume of group chromosome X-2 significantly decreases after the nuclei gain an extra chromosome 12.

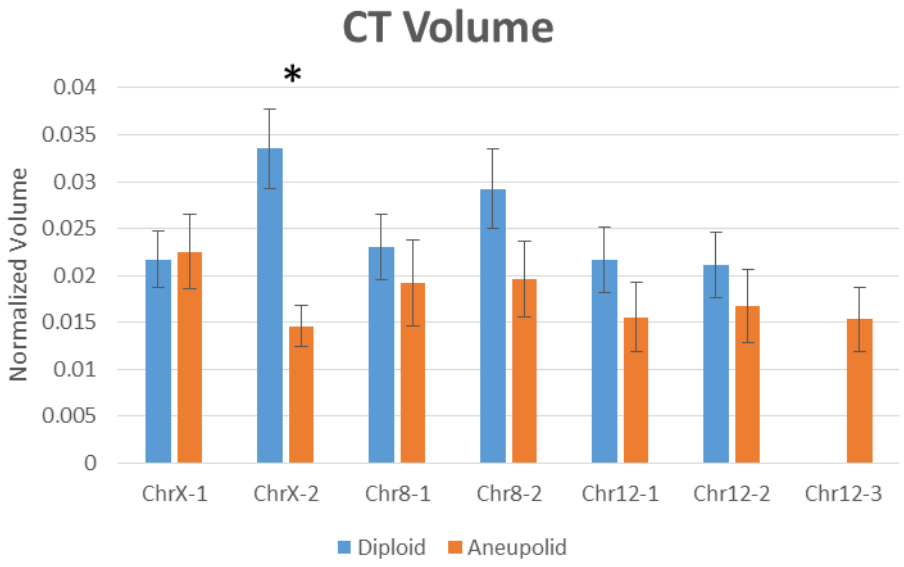

Figure 5.21. CT volume of chromosome X, 8, and 12 in normal and abnormal aneuploid cell samples.

The actual size of chromosome X, 8, and 12 are 155 Mbp, 145 Mbp, and 132 Mbp respectively at a descending order. The average CT volume from our results from diploid

group show a similar trend in size as the actual chromosome size (0.02761, 0.02615, and 0.02137 for chromosome X, 8, and 12 respectively).

5.3 Test with different local landmark

After the 3D model of CTs and nucleus are estimated with SPHARM shape descriptor, the next target for the formwork is to establish a coordinate system that is able to present positioning of chromosome territories within the nucleus with respect to a reference landmark. The process involves selecting local landmark (reference chromosome), labeling homologous chromosomes, and realignment as described in chapter 3. To test the robustness of the process that determines the specific 3D positioning of CTs, we select a different chromosome as the reference chromosome and compare the results. In the previous analysis chromosome X-1 was used as local landmark. Therefore in this analysis, we select chromosome 8 as the reference chromosome. The overall result on CT positioning should be similar when using different chromosomes as the reference chromosome.

The average position of the centroid of each chromosome in normal and abnormal sample groups in 3D space within nucleus with respect to chromosome 8-1 is shown in Figure 5.22. Same color markers are used to label the same homolog chromosome and the shapes of the marker indicate two different sample groups, dot for normal samples and triangle for abnormal samples. The dash lines circle the homolog chromosomes that statistically share similar space within the nucleus. As we can see in the figure, all the chromosomes from aneuploid sample group can find one homolog that occupy similar

position from the normal group except chromosome 12-2 in aneuploid. This is similar to the result when the relative position is computed with respect to chromosome X-1.

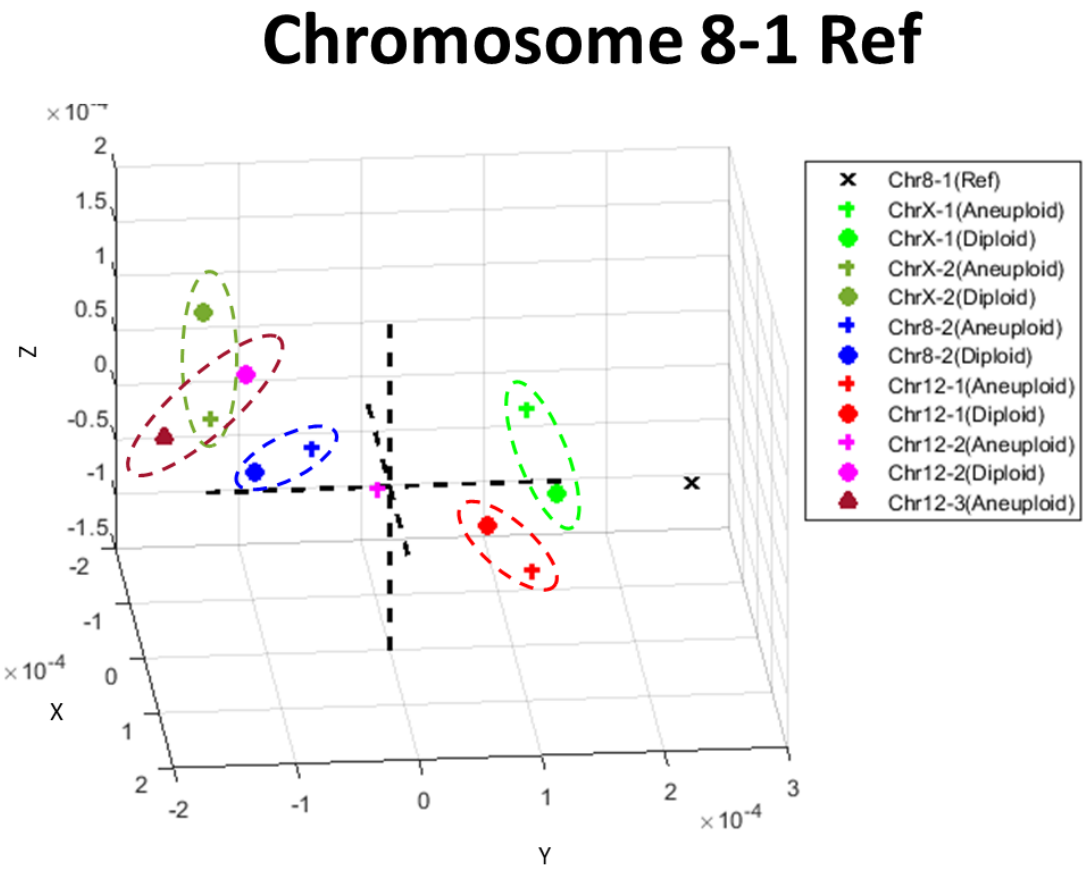

Figure 5.22. Average 3D position of mean centroid of each individual chromosome in normal diploid cells and aneuploid cells with respect to chromosome 8-1 as the local landmark.

Table 5.9 compares the p-value of chromosome X and 8 between normal and abnormal sample groups. The test is done with student's t-test with the same conditions as previous sections. As the table shows, the p-value for comparing to the chromosome that is labeled as the same homologous group in normal and abnormal samples (e.g., X-1 abnormal : X-1 normal) are much greater than 0.05, which indicates the equal mean null hypothesis is not rejected. This result would support the statement that chromosome territories is not randomly distributed. The p-value for different homologous chromosomes

(e.g., X-1 abnormal : X-2 normal) are all much smaller than 0.05. This means that each homolog occupies a unique territory. The positioning results for chromosome X and 8 with respect to chromosome X-1 as local landmark is matching to the results with chromosome 8-1 as the landmark.

Table 5.9. The p-value of student's t-test on the projection values of chromosome X and 8 between normal and abnormal sample groups respect to chromosome X or 8.

	Chromosome X			
Abnormal:Normal	X-1:X-1	X-1:X-2	X-2:X-1	X-2:X-2
X ref	1	1.51834E-12	2.95795E-05	0.17754134
8 ref	0.368346346	1.42946E-05	3.01539E-07	0.561331853
	Chromosome 8			
Abnormal:Normal	8-1:8-1	8-1:8-2	8-2:8-1	8-2:8-2
X ref	0.862157896	1.90867E-07	4.88977E-10	0.225677246
8 ref	1	3.96003E-10	1.94812E-14	0.752226259

We perform the same statistical tests as in the previous sections to the positioning result of chromosome 12 with respect to the new local landmark, chromosome 8-1, and the p-values of the tests are shown in Table 5.10. Comparing the result of p-values of the position with respect to the different local landmarks, we can find that regardless which landmark is selected, there are two of homologous of chromosome 12 that occupy similar space (p-value >> 0.05) when comparing the normal and abnormal sample group (highlighted in Table 5.10) and one homolog that is apart from the other two (chromosome 12-1 for X-1 as landmark and chromosome 12-2 for 8-1 as landmark). However, a different result occurs when using chromosome 8-1 as local landmark. The difference when comparing the position of chromosome 12-1 with respect to chromosome X-1 between diploid and aneuploid sample group is not statistically significant, but it is significant for the position of chromosome 12-2 respect to chromosome 8-1.

Table 5.10. The p-value of student's t-test on the projection values of chromosome 12 between normal and abnormal sample groups respect to chromosome X or 8.

	ChrX refrence			**Chr8 reference**		
	Chr12-1 Abnormal	Chr12-2 Abnormal	Chr12-3 Abnormal	Chr12-1 Abnormal	Chr12-2 Abnormal	Chr12-3 Abnormal
Chr12-1 Normal	0.040156	0.075773	1.421E-07	0.484512	0.003534	3.624E-12
Chr12-2 Normal	3.252E-11	0.001077	0.792140	1.10E-09	0.000736	0.145854
	ChrX refrence			**Chr8 refrence**		
Chr 12-1 vs 12-2 (Diploid, t-test)	6.26E-08			1.22071E-07		
(Aneuploid, ANOVA)	7.27E-10			7.19E-08		

As we can see in the Figure 5.22, all the chromosomes from aneuploid sample group can find one homolog that occupy similar position from the normal group except chromosome 12-2 in aneuploid. This is similar to the result of chromosome 12-1 when the relative position is respecting to chromosome X-1. Even though the chromosome label might be different, our framework shows a similar spatial organization of chromosome territories with different chromosome as local landmark.

5.4 Microarray Analysis

The hES cell samples at different passages (30, 37, 45, 50, 60, and 68) were analyzed with microarray (HumanWG-6 Expression BeadChip, Illumina, CA, USA). There were two technical replicates for each passage. Dr. Pati's lab provided microarray data that was pre-processed and completely analyzed. For quality control, GenomeStudio calculates a detection p-value that reports the confidence that a given transcript is expressed above the background defined by negative control probes. A transcript is called detected if the p-value is smaller than a user-defined threshold (0.001 is used for this study). The data were

preprocessed using the quantile normalization algorithm. Figure 5.23 shows the distribution before and after quantile normalization. This method assumes that the distribution of the expression values does not change dramatically between arrays. All arrays are adjusted so that they have an identical distribution.

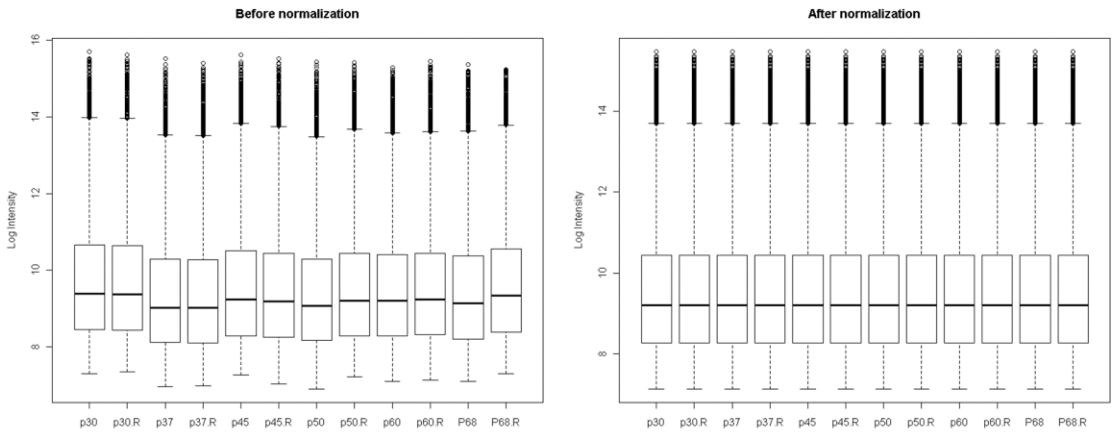

Figure 5.23. Box plot of raw (A) and normalized (B) fluorescence intensities showing the smallest and largest observations (whiskers), the median (black line) and the lower and upper quartile (box). Black circles represent outliers.

From a total of 48,803 probes only 9855 showed a detection p-value less than 0.001. To investigate the relationship between gene expression and CT organization, we focused on the overall gene expression through different passages of chromosome 8, 12, and X. There were 315, 510, and 388 genes detected from chromosome 8, 12, and X respectively. Some genes have more than one duplicates. The mean value of the gene expression intensity value of all duplicates of the same gene were used to represent that specific gene. As a result, we had 275 unique genes from chromosome 8, 428 from chromosome 12, and 321 from chromosome X. Figure 5.24 shows the average intensity value of genes from chromosome X, 8, and 12 across passage 30, 37, 45, 50, 60, 68 and their repeats. The overall gene expression of chromosome X and 8 remain unchanged throughout the passages and chromosome 12 showed a tendency of increasing over passages.

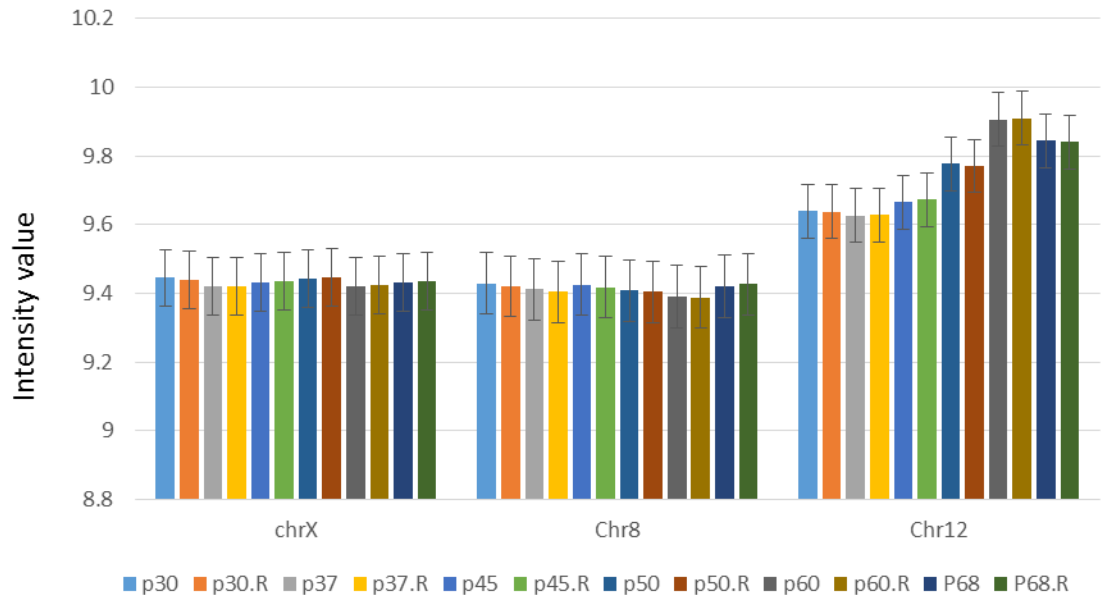

Figure 5.24. Gene expression of the genes that are distinct from the background and located in chromosome X, 8, and 12 across different passages and their repeats.

Paired t-test (two-tails, 95% confidence level) was performed to test if the difference in intensity value between certain passage and its repeat is significant, and no difference shown. We took the average of each passage and their repeat to represent the gene expression at particular passage (e.g., the average of gene expression of each gene at p30 and p30 repeat). To investigate if aneuploidy affects the gene expression, we conducted paired t-test on the gene expression at p68 and p30 since our sample set had grown about 70% aneuploidy by passage 68 and remained constant until the culture was terminated at passage 87. Table 5.11 shows the test result of the paired t-test on gene expression from chromosome X, 8, and 12. The p-value of the test from chromosome 12 is much less than 0.05 which indicates a significant difference between normal and aneuploid nuclei. Genes from the other two chromosomes did not show changes between the two passages. The test results of chromosome X and 8 were conserved with the CT positioning analysis that no

changes of chromosome X and 8 between normal and abnormal samples were observed in terms of the position relative to local landmark.

Table 5.11. Pared t-test on gene expression from chromosome X, 8, and 12 at passage 30 and passage68.

Chromosome X						
Passage	Mean	Variance	Observations	t Stat	P(T<=t)	t Critical
30	9.442173	2.21118	321	0.327454	0.743538	1.967405
68	9.434527	2.266153	321			
Chromosome 8						
Passage	Mean	Variance	Observations	t Stat	P(T<=t)	t Critical
30	9.424959	2.195503	275	0.042761	0.965923	1.96866
68	9.423933	2.224761	275			
Chromosome 12						
Passage	Mean	Variance	Observations	t Stat	P(T<=t)	t Critical
30	9.637894	2.62428	428	-11.814	4.62E-28	1.965535
68	9.842428	2.58616	428			

In addition to overall gene expression, we also selected the genes that significantly changed over time. The differential expression analysis was performed using a one-way ANOVA test for the time variable (passages) followed by post hoc analysis using the TukeyHSD pair test to determine which sample pairs show statistically significant differences [64]. The raw p values of the ANOVA test were adjusted using the Benjamini and Hochberg algorithm [65] for false discovery rate (FDR) control. The raw p-value for a gene is the probability of the observed fluorescence intensity or any more extreme intensities occurring, if the gene is not differentially expressed. FDR is the expected proportion of false positives among all significant genes as or more extreme than the one observed. 1270 genes showed differential expression for an FDR threshold of 0.001, including 32 genes from chromosome 8, 78 genes from chromosome 12, and 58 genes from chromosome X. For the genes that expressed differentially, genes in chromosome 12

increased with the development of aneuploidy whereas genes in chromosome 8 and X expressed less with growth of passages as shown in Figure 5.25.

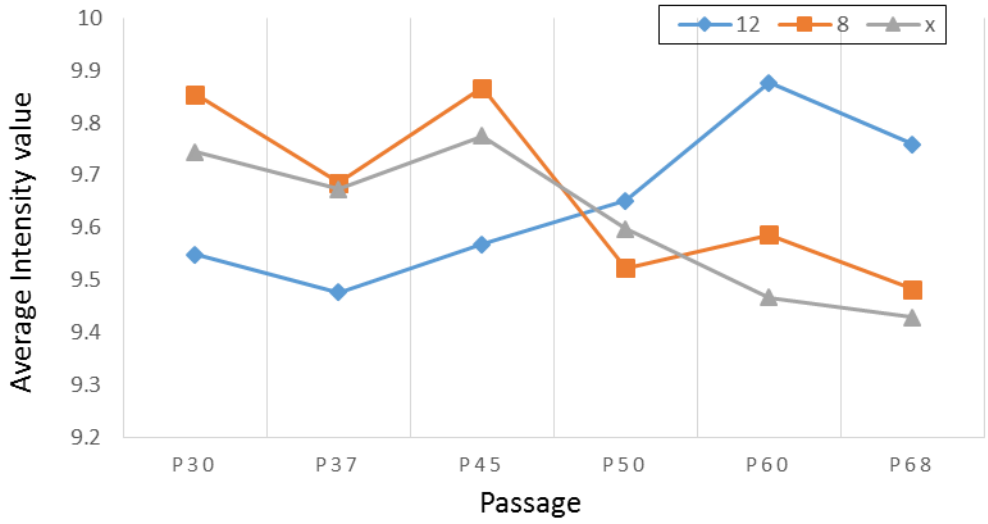

Figure 5.25. The average intensity of genes that showed differential expression in chromosome 8, 12, and X.

We listed out the genes in our target chromosomes that were always increasing/decreasing in terms of gene expression from one passage to the next with a tolerance for the opposite trend at a p value greater than 0.05 (no significant change from one passage to the next). The number of genes that were from chromosome 8, X and 12 that were always increasing are 3, 6 and 25 respectively. For always decreasing genes, there were 5 genes from chromosome 8, 13 from chromosome X, and 4 from chromosome 12. We utilized the database of gene co-regulation (dGCR) web tool[66] that analyzes the gene relationship according to gene expression over publicly available transcriptional data to identify co-regulated genes. We searched all the listed genes through dGCR to find the genes that are co-regulated and expressed increasingly/decreasingly throughout passages from our sample. Table 5.12 list the genes that were always increasing/decreasing in chromosome 8, 12, and X. The dGCR web tool computes the correlation between

individual genes and list the top 150 genes that have higher chance co-regulating to specific gene. We acquired the list for each gene that constantly changed and generated the correlation map of the genes that constantly changed according to the correlation score from dGCR web tool. A schemaball plot (Figure 5.26) was generated according to the co-regulated correlation among always increasing/decreasing genes in chromosome 8, X, and 12. Red color refers to genes from chromosome 12, green represents genes in chromosome X, and genes from chromosome 8 are in cyan. Genes that have a higher co-regulation score are linked with yellow lines and the intensity represents the score (the higher correlation, the brighter the linking line). Within the schemaball plot, we can identify the genes that have higher probability of co-regulation across inter- or intra- chromosomes. For example, RPL7, COX7B, and RPL41 are highly correlated. Figure 5.27 shows the gene expression throughout passages of genes that are co-regulated with gene(s) from other chromosome(s) that has a co-regulation correlation score from dGCR web tool less than -100. Color coded lines in red, green, and blue represent the location of the particular genes in chromosome X, 8, and 12 respectively.

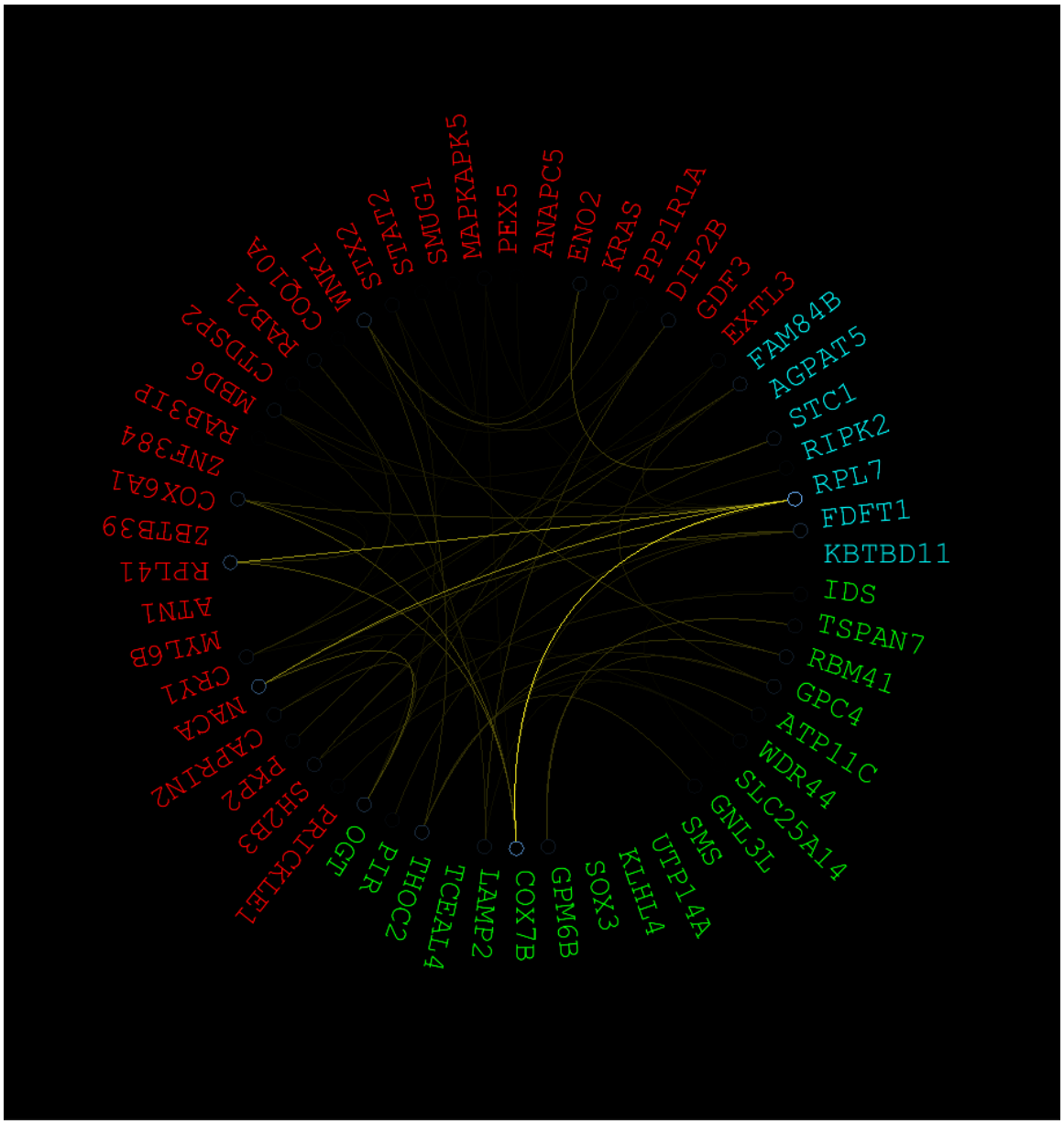

Figure 5.26. Schemaball representation of co-regulated always increasing/decreasing genes in chromosome 8, X, and 12.

Table 5.12. List of always increasing/decreasing genes in chromosome 8, 12, and X.

Chromosome 8		**Chromosome 12**		**Chromosome X**	
Always increase	Always decrease	Always increase	Always decrease	Always increase	Always decrease
KBTBD11	EXTL3	DIP2B	CAPRIN2	OGT	GPM6B
RPL7	AGPAT5	ENO2	PKP2	LAMP2	TSPAN7
FDFT1	FAM84B	PEX5	SH2B3	COX7B	IDS
	STC1	MAPKAPK5	PRICKLE1	PIR	SLC25A14
	RIPK2	SMUG1		THOC2	WDR44
		STAT2		TCEAL4	ATP11C
		CTDSP2			RBM41
		MBD6			GPC4
		RAB3IP			KLHL4
		ZNF384			UTP14A
		COX6A1			SMS
		RPL41			GNL3L
		CRY1			SOX3
		NACA			
		MYL6B			
		KRAS			
		WNK1			
		RAB21			
		STX2			
		GDF3			
		PPP1R1A			
		ANAPC5			
		COQ10A			
		ZBTB39			
		ATN1			

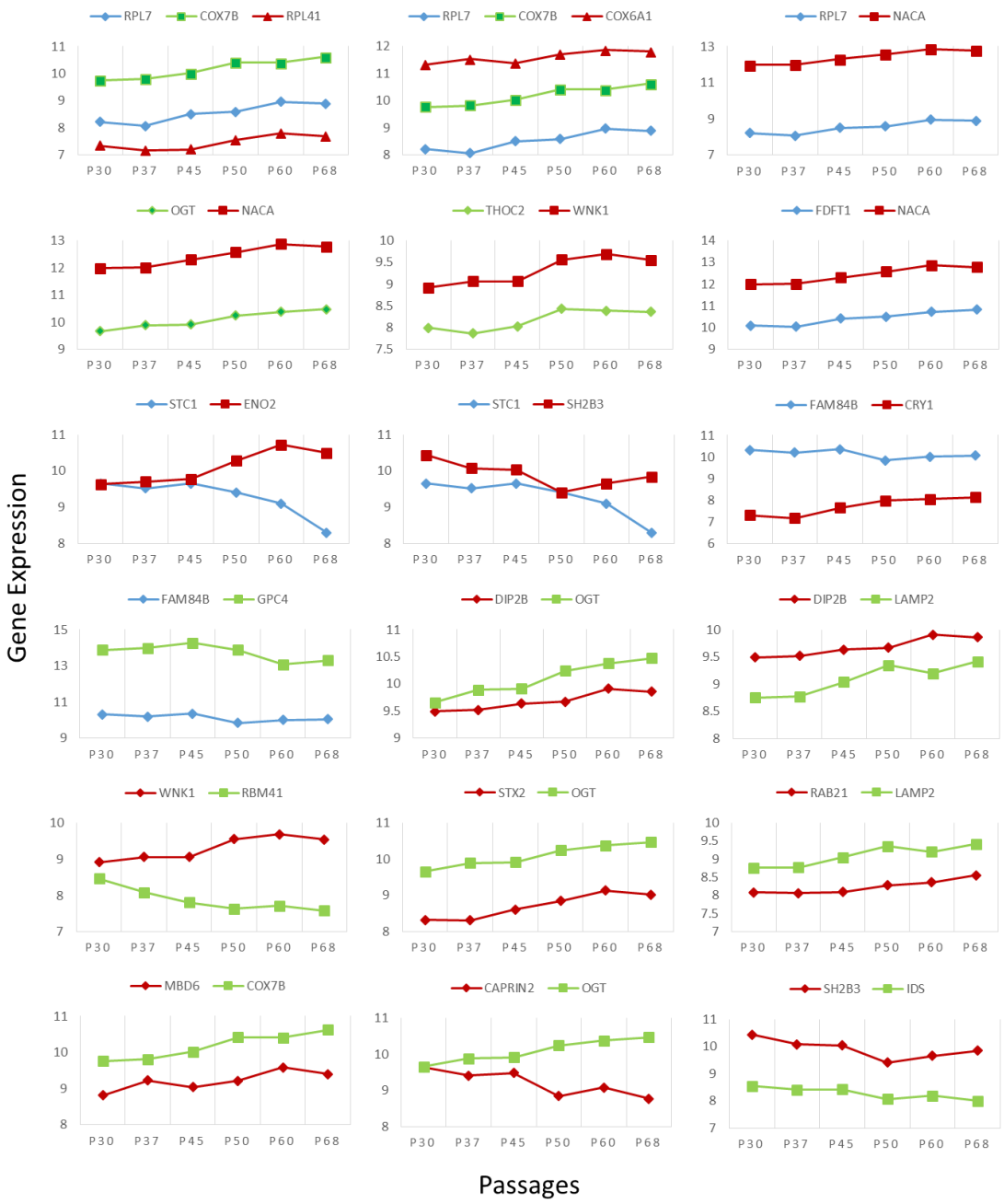

Figure 5.27. Gene expression at different passage for the genes that are co-regulated in chromosome X, 8, and 12.

Chapter 6 – Discussion

During interphase, chromosomes are in a less condensed form and each individual chromosome occupies a unique 3D space within the nucleus known as chromosome territories. It is known that the spatial organization of chromosome territories within nucleus is not randomly arranged for specific cell type. The development of confocal microscopy and florescent in situ hybridization with specific chromosome-painting probes facilitated the study of CTs at a higher order organization. Studies have reported that disruption of chromosomal organization could affect gene regulation, expression, and genomic instability and thus, may lead to unwanted genomic changes as seen in disease status such as cancer. Aneuploidy is frequently observed in cancer and alterations in gene expression mostly accompany with aneuploidy, but there are few studies that focus on the relationship between CT positioning and genome aberrations. These observations call attention to the importance of precisely determining the localization of CTs. Therefore, the goal of this study is to delineate disease mechanisms and understand how incorrectly expressed genes ultimately leaded to disease states with our 3D modeling tool that provides precise description of CT positioning.

We proposed a 3D modeling computational framework based on spherical harmonics shape descriptor (SPHARM) that accurately reconstructs the enclosed 3D surface of CTs and nuclei. SPHARM shape descriptor is able to provide the shape information that is invariant to scaling, translation, and rotation with proper alignment in the frequency domain. Due these proprieties, all the nuclear samples can be aligned to a common coordinate base on the shape of nucleus. The selection of global and local landmark made a finer alignment of CTs within the nucleus and allows the comparison of CT spatial

organization across different nuclear samples and further identify the characterization of chromosome territories within the 3D nucleus space.

We utilized human steam cells (hEs) WA09 (H9) that progressively developed aneuploidy over serial passaging. The cell line was provided by Dr. Pati, at the Baylor College of Medicine, at passage 73 and about 70% of the cells had acquired an extra chromosome 12 at passage 73. Thus, both normal (diploid) cells and abnormal (trisomy 12) samples were able to be acquired for the study. Multi-color 3D FISH with whole chromosome painting (WCP) probes and confocal microscopy were used to capture the chromosome topology within 3D nuclear volume. The WCP probes were targeting to chromosome 8, X, and 12 for a comprehensive analysis with two unchanged chromosomes (8 and X) as control reference and the aneuploidy on chromosome 12. The CT positioning of labeled chromosomes in the nucleus was performed using the proposed 3D modeling methods. The CTs were characterized in terms of their 3D position (mass centroid), radial distance, volume, inter-homologous distance, and intra-heterologous distance.

The CT spatial locations were represented in (x, y, z) coordinates relative to the reference chromosome class (chromosome 8 or X) and the CT positioning was analyzed accordingly. We demonstrated a robust alignment process and reference selection that result in a conserved CT spatial distribution while using different chromosomes as reference (chromosome 8 and X). The spatial location of CTs resulting from our analysis suggests that all the chromosome homologs of chromosome 8, X, and 12 occupy distinct territories which is consistent with previous studies [1], [63]. However, no significant disruption in terms of CT localization caused by aneuploid in chromosome 8, X, and 12 was observed which is conflicted to the previous studies that suggested the chromosomal

localization is altered during aneuploidy. Our results demonstrate that the diploid chromosome X homolog 2, chromosome 8 homolog 1 and 2 were found in the same position in both normal and aneuploid nuclei. Since the homolog labeling was based on the relative distance to the local landmark and the extra homolog in trisomy 12 could affect the distance ranking, the labeling result may be different between diploid and aneuploid samples, but both homologs in diploid samples still could find homolog that occupy the same position in aneuploid sample group. One homolog in trisomy 12 was observed located at a distinct position within the nucleus (i.e., significantly different in terms of position to other homologs in both normal and aneuploid sample group).

Radial distribution is commonly used for studying spatial organization of CTs. In our experiment, the radial distance of chromosome 8-2 significantly increased (p-value<0.05) during aneuploidy. Differences in radial distance also showed on the homologs in the same chromosome class and different pattern was observed between diploid and aneuploid group. The difference in terms of radial distance of the two homolog X's and 8's (i.e., X-1 vs X-2 and 8-1 vs 8-2) was not significant in normal sample group, but is significant in aneuploid sample group. Also, significance of difference in radial distance of homolog 12's was observed in normal group but not in the abnormal group.

The actual size of chromosome X, 8, and 12 are 155 Mbp, 145 Mbp, and 132 Mbp respectively. The average CT volume from our results from diploid group are 0.02761, 0.02615, and 0.02137 for chromosome X, 8, and 12 respectively. This show a similar trend where the size of chromosome X > chromosome 8 > chromosome 12. One important finding was the chromosome volume of chromosome 8-2 decreased significantly in aneuploidy group.

The mean intra-homologous distance from our results are constant between normal and abnormal sample groups in chromosome 8 and chromosome X whereas, the homologous distance of chromosome 12 is significantly decreased when gaining one extra chromosome 12. Although trisomy 12 does not affect the intra-homologous separation of homolog 8's and X's, our results shows significant decreases in the inter-heterologous distances of chromosome 8 to 12, X to 12, and 8 to X. The distances between CTs are critical in development of cancer. Studies have shown the closer proximity have a greater frequencies of translocation [38], [67].

Our results demonstrate that the overall gene expression of chromosome 8, X does not change with the development of aneuploidy, but for chromosome 12 we observe an overall increase in gene expression level as expected due to the presence of an extra chromosome 12. These observations are consistent with our results of CT organization which indicates significant radial position changes in chromosome 12 in aneuploid nuclei. The genes in chromosome 8 and X that were expressed differentially decreased in expression level over passages, whereas the ones in chromosome 12 increased in terms of gene expression level.

This dissertation demonstrated a novel approach to model the chromosome territories in 3D nuclear space and quantitatively analyze spatial organization of chromosome territories of diploid verses aneuploid genomes. I have shown experimental evidence supporting that changes in ploidy are refelective with changes in spatial organization of chromosome territories and gene expression. The modeling framework is able to precisely reconstruct the shape of CTs and nuclei from 3D FISH signal and provide a cross sample evaluation in terms of size and position in a common (x, y, z) coordinate.

In the future, SPHARM based surface modeling method brings the possibility of investigation in the morphometric analysis of nuclei and/or chromosome territories. The SPHARM coefficients has potential for characterizing and comparing nuclear shape, which might enable the model to be part of the cancer diagnostic process since the cancer nuclei are generally larger and have more irregular shape than normal nuclei. The study could also be extended by cooperating more chromosome probes and/or more cell lines for a complete evaluation of the relationship between CT spatial organization and genotype and ultimately being able to distinguish disease cells from non-disease cells. Further, if changes in specific genes expression patterns are identified, using FISH probes for the target genes can also be performed to map any changes in loci positioning using the 3D modeling tools developed.

The complexity of how disease changes influence CT organization is further compounded due to cell-to-cell variability, tissue specificity, and the heterotypic karyotypes associated with different disease states. Therefore, the study of nuclear organization still is still it its infancy. There is a clear need of further investigations in chromosomal topography. The proposed methods and experimental results may contribute to the journey of unraveling the understanding of the underlying mechanisms of gene regulation, genome instabilities and rearrangements.

Bibliography

[1] T. Cremer and M. Cremer, “Chromosome territories.,” *Cold Spring Harb. Perspect. Biol.*, vol. 2, no. 3, pp. 1–22, 2010.

[2] J. A. Croft, J. M. Bridger, S. Boyle, P. Perry, P. Teague, and W. A. Bickmore, “Differences in the localization and morphology of chromosomes in the human nucleus,” *J. Cell Biol.*, vol. 145, no. 6, pp. 1119–1131, 1999.

[3] S. Boyle, S. Gilchrist, J. M. Bridger, N. L. Mahy, J. A. Ellis, and W. A. Bickmore, “The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells,” vol. 10, no. 3, pp. 211–220, 2001.

[4] H. Bin Sun, J. Shen, and H. Yokota, “Size-dependent positioning of human chromosomes in interphase nuclei,” *Biophys. J.*, vol. 79, no. 1, pp. 184–190, 2000.

[5] M. Cremer, J. V. Hase, T. Volm, A. Brero, G. Kreth, J. Walter, C. Fischer, I. Solovei, C. Cremer, and T. Cremer, “Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells,” *Chromosom. Res.*, vol. 9, no. 7, pp. 541–567, 2001.

[6] A. Bolzer, G. Kreth, I. Solovei, D. Koehler, K. Saracoglu, C. Fauth, S. Müller, R. Eils, C. Cremer, M. R. Speicher, and T. Cremer, “Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes,” *PLoS Biol.*, vol. 3, no. 5, pp. 0826–0842, 2005.

[7] M. Cremer, F. Grasser, C. Lanctôt, S. Müller, M. Neusser, R. Zinner, I. Solovei, and T. Cremer, “Multicolor 3D fluorescence in situ hybridization for imaging interphase chromosomes,” *Methods Mol. Biol.*, vol. 463, pp. 205–239, 2008.

[8] M. Cremer, K. Küpper, B. Wagler, L. Wizelman, J. V. Hase, Y. Weiland, L. Kreja,

J. Diebold, M. R. Speicher, and T. Cremer, “Inheritance of gene density-related higher order chromatin arrangements in normal and tumor cell nuclei,” *J. Cell Biol.*, vol. 162, no. 5, pp. 809–820, 2003.

[9] K. J. Meaburn, T. Misteli, and E. Soutoglou, “Spatial genome organization in the formation of chromosomal,” vol. 17, no. 1, pp. 80–90, 2007.

[10] L. E. Finlan, D. Sproul, I. Thomson, S. Boyle, E. Kerr, P. Perry, B. Ylstra, J. R. Chubb, and W. A. Bickmore, “Recruitment to the nuclear periphery can alter expression of genes in human cells,” *PLoS Genet.*, vol. 4, no. 3, 2008.

[11] L. A. Parada, P. G. McQueen, P. J. Munson, and T. Misteli, “Conservation of relative chromosome positioning in normal and cancer cells,” *Curr. Biol.*, vol. 12, no. 19, pp. 1692–1697, 2002.

[12] J. J. Roix, P. G. McQueen, P. J. Munson, L. A. Parada, and T. Misteli, “Spatial proximity of translocation-prone gene loci in human lymphomas,” *Nat. Genet.*, vol. 34, no. 3, pp. 287–291, 2003.

[13] A. E. Schröck, S. Manoir, T. Veldman, B. Schoell, J. Wienberg, Y. Ning, D. H. Ledbetter, D. Soenksen, Y. Garini, and T. Ried, “Multicolor Spectral Karyotyping of Human Chromosomes Published by : American Association for the Advancement of Science Stable URL : https://www.jstor.org/stable/2890516 digitize , preserve and extend access to Science Multicolor Spectral Karyotyping of H,” vol. 273, no. 5274, pp. 494–497, 2019.

[14] T. Cremer, C. Cremer, H. Baumann, E. K. Luedtke, K. Sperling, V. Teuber, and C. Zorn, “Rabl’s model of the interphase chromosome arrangement tested in Chinise hamster cells by premature chromosome condensation and laser-UV-microbeam

experiments," *Hum. Genet.*, vol. 60, no. 1, pp. 46–56, 1982.

[15] T. Cremer, C. Cremer, T. Schneider, H. Baumann, L. Hens, and M. Kirsch-Volders, "Analysis of chromosome positions in the interphase nucleus of Chinese hamster cells by laser-UV-microirradiation experiments," *Hum. Genet.*, vol. 62, no. 3, pp. 201–209, 1982.

[16] S. M. Stack, D. B. Brown, and W. C. Dewey, "Visualization of interphase chromosomes," *J. Cell Sci.*, vol. 26, no. 1, p. 281 LP-299, Aug. 1977.

[17] T. Misteli, "Chromosome territories: The arrangement of chromosomes in the nucleus.," *Nat. Educ.*, vol. 1, no. 1, 2008.

[18] K. Küpper, A. Kölbl, D. Biener, S. Dittrich, and T. Thormeyer, "Radial chromatin positioning is shaped by local gene density , not by gene expression," vol. 116, no. 3, pp. 285–306, 2009.

[19] R. Mayer, A. Brero, J. Von Hase, T. Schroeder, T. Cremer, and S. Dietzel, "Common themes and cell type specific variations of higher order chromatin arrangements in the mouse," vol. 22, pp. 1–22, 2005.

[20] C. Federico, C. Scavo, C. Daniela, C. Salvatore, S. Saccone, and G. Bernardi, "Gene-rich and gene-poor chromosomal regions have different locations in the interphase nuclei of cold-blooded vertebrates," pp. 123–128, 2006.

[21] S. Goetze, J. Mateos-langerak, H. J. Gierman, W. De Leeuw, O. Giromus, M. H. G. Indemans, J. Koster, V. Ondrej, R. Versteeg, and R. Van Driel, "The Three-Dimensional Structure of Human Interphase Chromosomes Is Related to the Transcriptome Map," vol. 27, no. 12, pp. 4475–4487, 2007.

[22] J. Dixon, D. Gorkin, B. R.-M. cell, and undefined 2016, "Chromatin domains: the

unit of chromosome organization," *Elsevier*, vol. 62, no. 5, pp. 668–680, 2017.

[23] R. B. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, "Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions," *Biotechnol. Bioeng.*, vol. 50, no. 5, pp. 294–306, 2010.

[24] D. A. Jackson and A. Pombo, "Replicon clusters are stable units of chromosome structure: Evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells," *J. Cell Biol.*, vol. 140, no. 6, pp. 1285–1295, 1998.

[25] M. Cremer, V. J. Schmid, F. Kraus, Y. Markaki, I. Hellmann, A. Maiser, H. Leonhardt, S. John, J. Stamatoyannopoulos, and T. Cremer, "Initial high‐resolution microscopic mapping of active and inactive regulatory sequences proves non‐random 3D arrangements in chromatin domain clusters," *Epigenetics Chromatin*, pp. 1–17, 2017.

[26] S. D. Zhao R, Bodnar M, "Nuclear neighborhoods and gene expression," *Mol. Cell. Biol.*, vol. 19, no. 2, pp. 172–179, 2010.

[27] P. Oberdoerffer and D. A. Sinclair, "The role of nuclear architecture in genomic instability and ageing," *Nat. Rev. Mol. Cell Biol.*, vol. 8, p. 692, Sep. 2007.

[28] H. Tanabe, F. A. Habermann, I. Solovei, M. Cremer, and T. Cremer, "Non-random radial arrangements of interphase chromosome territories: Evolutionary considerations and functional implications," *Mutat. Res. - Fundam. Mol. Mech. Mutagen.*, vol. 504, no. 1–2, pp. 37–45, 2002.

[29] A. R. Leitch, "Higher Levels of Organization in the Interphase Nucleus of Cycling and Differentiated Cells Higher Levels of Organization in the Interphase Nucleus of

Cycling and Differentiated Cells," *Microbiol. Mol. Biol. Rev.*, vol. 64, no. 1, pp. 138–52, 2000.

[30] M. R. Branco and A. Pombo, "Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations," *PLoS Biol.*, vol. 4, no. 5, pp. 780–788, 2006.

[31] S. Stadler, V. Schnapp, R. Mayer, S. Stein, C. Cremer, C. Bonifer, T. Cremer, and S. Dietzel, "The architecture of chicken chromosome territories changes during differentiation," *BMC Cell Biol.*, vol. 5, pp. 1–17, 2004.

[32] I. Solovei, M. Kreysing, C. Lanctôt, S. Kösem, L. Peichl, T. Cremer, J. Guck, and B. Joffe, "Nuclear Architecture of Rod Photoreceptor Cells Adapts to Vision in Mammalian Evolution," *Cell*, vol. 137, no. 2, pp. 356–368, 2009.

[33] G. Kreth, J. Finsterle, J. Von Hase, M. Cremer, and C. Cremer, "Radial Arrangement of Chromosome Territories in Human Cell Nuclei: A Computer Model Approach Based on Gene Density Indicates a Probabilistic Global Positioning Code," *Biophys. J.*, vol. 86, no. 5, pp. 2803–2812, 2004.

[34] A. Shete, P. ; Rao, D. Pati, and F. Merchant, "Spatial Quantitation of FISH signals in Diploid versus Aneuploid Nuclei," *Cytom. Part A J. Int. Soc. Anal. Cytol.*, vol. 49, no. 18, pp. 1841–1850, 2009.

[35] T. Misteli, "Spatial positioning: A new dimension in genome function," *Cell*, vol. 119, no. 2, pp. 153–156, 2004.

[36] T. Misteli, "Beyond the Sequence: Cellular Organization of Genome Function," *Cell*, vol. 128, no. 4, pp. 787–800, 2007.

[37] T. Misteli, "Higher-order genome organization in human disease," *Cold Spring*

Harb. Perspect. Biol., vol. 2, no. 8, p. a000794, 2010.

[38] L. Brianna Caddle, J. L. Grant, J. Szatkiewicz, J. Van Hase, B. J. Shirley, J. Bewersdorf, C. Cremer, A. Arneodo, A. Khalil, and K. D. Mills, "Chromosome neighborhood composition determines translocation outcomes after exposure to high-dose radiation in primary cells," *Chromosom. Res.*, vol. 15, no. 8, pp. 1061–1073, 2007.

[39] A. B. Berger, G. G. Cabal, E. Fabre, T. Duong, H. Buc, U. Nehrbass, J. C. Olivo-Marin, O. Gadal, and C. Zimmer, "High-resolution statistical mapping reveals gene territories in live yeast," *Nat. Methods*, vol. 5, no. 12, pp. 1031–1037, 2008.

[40] A. Khalil, J. L. Grant, L. B. Caddle, E. Atzema, K. D. Mills, and A. Arneodo, "Chromosome territories have a highly nonspherical morphology and nonrandom positioning," *Chromosom. Res.*, vol. 15, no. 7, pp. 899–916, 2007.

[41] T. Misteli, "Concepts in nuclear architecture," *BioEssays*, vol. 27, no. 5, pp. 477–487, 2005.

[42] Carmelinda Ruggiero, E. Jeffrey Metter, and Antonio Cherubini, "White blood cell count and mortality in the Baltimore Longitudinal Study of Aging," *J. Am. Coll. Cardiol.*, vol. 49, no. 18, pp. 1841–1850, 2007.

[43] M. J. Zeitz, L. Mukherjee, S. Bhattacharya, J. Xu, and R. Berezney, "A Probabilistic Model for the Arrangement of a Subset of Human Chromosome Territories in WI38 Human Fibroblasts," no. June, pp. 120–129, 2009.

[44] S. Kemeny, C. Tatout, G. Salaun, C. Pebrel-richard, C. Goumy, N. Ollier, E. Maurin, B. Pereira, P. Vago, and L. Gouas, "Spatial organization of chromosome territories in the interphase nucleus of trisomy 21 cells," pp. 247–259, 2018.

[45] N. Pavelka, G. Rancati, J. Zhu, W. D. Bradford, A. Saraf, L. Florens, B. W. Sanderson, G. L. Hattem, and R. Li, "Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast," *Nature*, vol. 468, no. 7321, pp. 321–325, 2011.

[46] K. Sengupta, M. B. Upender, L. Barenboim-Stapleton, Q. T. Nguyen, S. M. Wincovitch, S. H. Garfield, M. J. Difilippantonio, and T. Ried, "Artificially introduced aneuploid chromosomes assume a conserved position in colon cancer cells," *PLoS One*, vol. 2, no. 2, 2007.

[47] H. Huang, L. Shen, R. Zhang, F. Makedon, B. Hettleman, and J. Pearlman, "Surface alignment of 3D spherical harmonic models: application to cardiac MRI analysis.," *Med. Image Comput. Comput. Assist. Interv.*, vol. 8, no. Pt 1, pp. 67–74, 2005.

[48] G. Gerig and M. Styner, "Shape analysis of brain ventricles using SPHARM," pp. 171–178, 2001.

[49] F. P. Kuhl and C. R. Giardina, "Elliptic Fourier features of a closed contour," *Comput. Graph. Image Process.*, vol. 18, pp. 236–258, 1982.

[50] L. Shen, H. Farid, and M. A. McPeek, "Modeling three-dimensional morphological structures using spherical harmonics," *Evolution (N. Y).*, vol. 63, no. 4, pp. 1003–1016, 2009.

[51] C. Brechbühler, G. Gerig, and O. Kübler, "Parametrization of Closed Surfaces for 3-D Shape Description," *Comput. Vis. Image Underst.*, vol. 61, no. 2, pp. 154–170, 1995.

[52] R. Green, "S Pherical Harmonic Lighting," 2003.

[53] A. E. Wiblin, "Distinctive nuclear organisation of centromeres and regions involved

in pluripotency in human embryonic stem cells," *J. Cell Sci.*, vol. 118, no. 17, pp. 3861–3868, 2005.

[54] E. Arias-Castro and D. L. Donoho, "Does median filtering truly preserve edges better than linear filtering?," *Ann. Stat.*, vol. 37, no. 3, pp. 1172–1206, 2009.

[55] L. Shen, "SPHARM-MAT Documentation," 2010.

[56] M. Styner, I. Oguz, S. Xu, C. Brechbuhler, D. Pantazis, J. J. Levitt, M. E. Shenton, and G. Gerig, "Framework for the Statistical Shape Analysis of Brain Structures using SPHARM-PDM," *Insight J*, no. 1071, pp. 242–250, 2006.

[57] L. Shen and F. Makedon, "Spherical mapping for processing of 3D closed surfaces," *Image Vis. Comput.*, vol. 24, no. 7, pp. 743–761, 2006.

[58] C. Brechbühler, G. Gerig, and O. Kübler, "Parametrization of Closed Surfaces for 3-D Shape Description," *Comput. Vis. Image Underst.*, vol. 61, no. 2, pp. 154–170, Mar. 1995.

[59] S. Althloothi, M. H. Mahoor, R. M. Voyles, and S. Member, "A Robust Method for Rotation Estimation Using Spherical Harmonics Representation," *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2306–2316, 2013.

[60] L. Shen, H. Huang, F. Makedon, and A. J. Saykin, "Efficient registration of 3D SPHARM surfaces," *Proc. - Fourth Can. Conf. Comput. Robot Vision, CRV 2007*, pp. 81–88, 2007.

[61] T. Mm Oller and B. Trumbore, "Fast, Minimum Storage Ray/Triangle Intersection," *Acm Siggraph*, no. 1, pp. 1–7, 2005.

[62] G. Hoffmann, "Sphere Tessellation by Icosahedron Subdivision." [Online]. Available: http://www.docs-hoffmann.de/ikos27042002.pdf.

[63] A. Shete, P. Rao, D. Pati, and F. Merchant, “Spatial Quantitation of FISH Signals in Diploid Versus Aneuploid Nuclei,” no. 6, 2014.

[64] T. R. C. Team, “R: A Language and Environment for Statistical Computing,” 2019. [Online]. Available: https://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf.

[65] Y. Benjamini, Y. Hochberg, and H. Y. Benjamini, Yoav, “Controlling the false discovery rate: a practical and powerful approach to multiple testing.,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 57, no. 1. pp. 289–300, 1995.

[66] G. Williams, “Database of Gene Co-Regulation (dGCR): A Web Tool for Analysing Patterns of Gene Co-regulation across Publicly Available Expression Data,” *J. Genomics*, vol. 3, pp. 29–35, 2015.

[67] N. V. Marella, S. Bhattacharya, L. Mukherjee, J. Xu, and R. Berezney, “Cell type specific chromosome territory organization in the interphase nucleus of normal and cancer cells,” *J. Cell. Physiol.*, vol. 221, no. 1, pp. 130–138, 2009.