

Ruprecht-Karls-Universität Heidelberg
Naturwissenschaftlich-Mathematische Gesamtfakultät

Quantitative Analysis of Chromosome/Gene Spatial Distribution

Dissertation
Submitted by Master of Science Juntao Gao
Heidelberg, February 2005

Dissertation

submitted to the
combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by

Master of Science: Juntao Gao
Born in: HeNan, P.R.China

Oral examination:

Quantitative Analysis of Chromosome/Gene Spatial Distribution

Referees: Prof. Dr. Roland Eils
Prof. Dr. Hans Peter Meinzer

To my parents for their life-long love and endless encouragement

Zusammenfassung

Die Fortschritte in den Bereichen Zellbildung und Mikroskopie haben eine intensive Untersuchung der räumlichen Anordnung von Chromosomen bzw. Genen im Verlauf der letzten Jahre möglich gemacht. Es wurden bereits Algorithmen zur Quantifizierung der räumlichen Anordnung von Chromosomen bzw. Genen entwickelt, wobei die meisten dieser Methoden jedoch auf zweidimensionalen (2D) Bilddaten basieren. Um dreidimensionale (3D) konfokale Bilddaten verarbeiten zu können, ist es notwendig neue Algorithmen zu entwickeln, die auf 3D Datensätzen basieren.

In dieser Arbeit werden neue Methoden zur Beschreibung, Analyse und Visualisierung der 3D Verteilung von Chromosomen und Genen in fixierten Zellkernen in 3D Bilddaten präsentiert, die basierend auf Konzepten der objektorientierten Programmierung in der Programmiersprache Java implementiert wurden. Kapitel 2 beschreibt verschiedene Softwarewerkzeuge zur Bestimmung von Ähnlichkeiten der Anordnung von Chromosomen unter Verwendung der Krümmungsenergie von Thin-plate Splines sowie zur Berechnung von geometrischen Mittelpunkten, Distanzen und Winkeln. Zwei anwendungsorientierte Projekte zur räumlichen Verteilung von Chromosomen bzw. Genen werden in Kapitel 3 bzw. Kapitel 4 vorgestellt. Die Vorteile, Grenzen und weitere Verbesserungen von diesen Computermethoden werden ausführlich in Kapitel 5 behandelt. Danach wird ein neues Modell der Vererbung der räumlichen Chromosomenordnung erläutert.

Die quantitative Analyse in Kapitel 3 zeigt, dass die Unterschiede in der Anordnung von Chromosomenterritorien kontinuierlich mit der Anzahl der Zellgenerationen (d.h. Zellteilungen) zunehmen. In HeLa Zellklonen sind die Unähnlichkeiten in der Anordnung von Chromosomenterritorien nach fünf oder sechs Zellteilungen bereits so groß wie die zwischen unverwandten, zufällig ausgewählten Zellen. Die quantitative Analyse in Kapitel 4 zeigt, dass während der Interphase die Positionen der untersuchten Gene (*MLL* und fünf seiner Translokationspartner) sowie von vier chromosomalen Kontroll-Loci ein charakteristisches Verteilungsmuster innerhalb des Zellkerns besitzen. Dies gilt für jede der untersuchten hämatopoietischen Zellen.

Die in dieser Arbeit präsentierten Methoden zur Bestimmung der Ähnlichkeit der Anordnung von Chromosomen unter Verwendung von punktbasierter Registrierung liefern zum ersten Mal Beispiele zur Analyse und Bewertung der Vererbung einer räumlichen Verteilung von Chromosomen über mehrere Zellteilungen hinweg. Gleichzeitig wird zum ersten Mal die räumliche Verteilung von Genen, vor allem des Genes *MLL* und einiger Translokationspartner, im 3D Raum des Interphasezellkerns hämatopoietischer Zellen quantitativ beschrieben.

Abstract

The spatial arrangement of chromosomes/genes has been studied intensively over the last several years facilitated by the advances in cell imaging technology and microscopy. Although algorithms have been developed to quantify the spatial arrangement of chromosomes/genes, most of these methods use two-dimensional image data. To obtain information from three dimensional (3D) confocal image data, it is necessary to develop new computational tools based on 3D data.

New 3D computational tools are developed using the concept of object-oriented programming and Java programming language, in order to describe, analyze and visualize the 3D distribution of chromosomes and genes in fixed nuclei. Computational tools to determine the similarity using the bending energy of thin-plate splines and for the calculation of the geometric center, distance and angle of chromosomes/genes are presented in chapter 2. Two application-oriented projects about chromosome/gene spatial distribution are presented in chapter 3 and chapter 4, respectively. The advantages, limits and further improvement of these computational tools are discussed in chapter 5, and one new model about the inherity of chromosome spatial distribution is presented thereafter.

The quantitative analysis in chapter 3 reveals that the dissimilarities in chromosome territory arrangements increase monotonously during different cell divisions. In HeLa cell clones, dissimilarities among cells in every clone reach the level of randomly chosen cells after five or six cell divisions. In chapter 4 it is shown that in all studied hematopoietic cells, the localization of gene *MLL*, its five translocation partners and four chromosomal control loci possesses a characteristic distribution pattern in the interphase nucleus.

For the first time, the methods presented in this thesis provide examples in which computing the similarity based on point set registration allows the analysis and evaluation of inheritance of chromosome spatial distribution during several cell divisions. And it is also the first time that the spatial distribution of genes, with focus on gene *MLL* and some of its translocation partners, was analysed and quantified within the 3D space of interphase nuclei of hematopoietic cells.

Contents

1. Chapter 1 Introduction	1
1.1 Translocation and spatial positioning	1
1.1.1 Cell nucleus and chromosomes	1
1.1.2 Cancer and translocation	3
1.1.3 Translocation and gene position	4
1.1.4 Gene position, proximity effects and chromosome position	7
1.1.5 Spatial positioning and clinical medicine	8
1.2 Chromosome positioning	9
1.2.1 Why is chromosome positioning important?	9
1.2.1.1 Chromosome positioning and gene expression	9
1.2.1.2 Gene position near heterochromatin or in nucleolus	9
1.2.2 The history of chromosome positioning research	10
1.2.3 The tethering of interphase chromosomes	12
1.2.4 Why is chromosome positioning research difficult?	14
1.3 Image processing, computer vision and the applications	15
1.3.1 Image processing and computer vision	15
1.3.2 Image processing applications in biological data	17
1.3.2.1 Analyzing the dynamics of cell image data	17
1.3.2.2 Quantitative analysis and visualization of biological image data	18
1.3.2.3 Development of mathematical model	18
1.3.3 How to apply image processing tools into chromosome/gene position data?	19
1.3.4 What is new in this thesis	20
2. Chapter 2 Computational methods for quantitative analysis and visualization of nuclear structure in fixed cell nuclei	23
2.1 The image analysis pipeline used in this work	23
2.2 Landmark data	25
2.2.1 Geometric center of spatial entity	25
2.2.2 Landmarks – definition	26
2.2.3 Configuration space	27

2.3 Analysis of landmark data by the means of angles and distances	27
2.3.1 Size measure	27
2.3.2 The Hausdorff distance	28
2.3.3 Normalization of the configuration	30
2.3.4 Distance Difference (DD) of two homologues in the same chromosome of two cells	31
2.3.5 Angle calculation	33
2.4 Analysis of landmark data by the means of geometric transformations	35
2.4.1 Rigid and affine transformation	35
2.4.2 Thin-plate splines defining planar or volume transformations	37
2.4.3 Measuring the similarity of shapes and spatial configurations based on the integral bending norm	39
2.5 Quantitative analysis and visualization of cell nuclear structures	40
2.5.1 Adaptation of the landmark based methods to the investigated data	40
2.5.2 Statistical analysis	41
2.5.3 Visualization	41
2.5.4 Aspects of implementation	42
2.5.5 Principles of Object-Oriented Programming (OOP)	43
3. Chapter 3 Quantitative analysis of the inheritance and changes of chromosome arrangement in proliferating human cells	44
3.1 introduction	44
3.1.1 The chromosome position similarity in daughter nuclei	44
3.1.2 Recent studies of chromosome arrangement inheritance	45
3.1.3 New strategy to study changes in chromosome arrangement between different cell generations	46
3.2 Data description	47
3.2.1 Cell culture and the confocal images	47
3.2.2 Different cell generations	49
3.2.3 Data description of the four cell lines	50
3.2.4 RCC clone definition and its realization	51
3.3 Quantification of the similarity of CT spatial arrangement	52

3.3.1 BE calculation depends on the setting of landmarks	52
3.3.2 The landmark setting in four cell lines	53
3.3.2.1 Cell combinations in every cell clone	54
3.3.2.2 Homologue combinations in every two cells	54
3.3.2.3 The calculation of BE	55
3.3.3 Codes improvement to calculate the BE	56
3.3.4 PCA-normalized spheres of DLD cell line	56
3.4 BE and DD results	57
3.4.1 BE result in several cell lines	57
3.4.1.1 HeLa cell line	57
3.4.1.2 HMEC cell line	58
3.4.1.3 HFb_CB cell line	58
3.4.2 The identification of sister and cousin cell pairs in four-cell clones	58
3.4.3 DD result in every cell line	61
3.4.3.1 HeLa cell line	61
3.4.3.2 HMEC cell line	62
3.4.3.3 HFb_CB cell line	63
3.4.3.4 DLD cell line	63
3.4.4 The comparison of DD and BE approaches	64
3.5 Conclusion	64
4. Chapter 4 Quantitative analysis of the spatial distribution of gene <i>MLL</i> and its translocation partners	66
4.1 The <i>MLL</i> gene and some of its translocation partners	66
4.1.1 The introduction to gene <i>MLL</i>	66
4.1.2 Some translocation partners of gene <i>MLL</i> (<i>AF9</i> , <i>AF4</i> , <i>AF6</i> , <i>ENL</i> , <i>ELL</i>)	67
4.2 Data description	68
4.2.1 Cell culture, cell cloning, 3D-FISH and surface reconstruction	69
4.2.2 Cell lines used in this study	69
4.3 Result and conclusion	70
4.3.1 The distance between the nuclear center (NC) and genes	70
4.3.2 Distance among genes in hematopoietic cells	70
4.3.3 The shortest distance between genes and nuclear surface	70

4.3.4 Angles among genes	72
5. Chapter 5 Discussion and future work	74
5.1 Discussion of the proposed computational tools	74
5.1.1 Geometric center and the center of gravity	74
5.1.1.1 Is the geometric center method a crude method?	74
5.1.1.2 Geometric center is different from center of gravity	75
5.1.2 The advantage of these tools	75
5.1.2.1 Use of the Bending Energy is one possibility to evaluate the similarity of chromosome spatial distribution during different cell divisions	76
5.1.2.2 Error checking in the original data	77
5.1.3 Improvements of these tools	78
5.1.3.1 Triangulation: Marching cube algorithm	78
5.1.3.2 Segmentation	79
5.2 Models for chromosome positioning during interphase and mitosis	81
5.2.1 Models for interphase chromosome positioning	81
5.2.2 Models for mitotic chromosome positioning	84
5.2.3 A new model for chromosome position inherity during cell divisions	87
5.3 Future direction: to develop new quantitative methods for chromatin spatial distribution at three levels	88
5.3.1 The integration of spatial organization quantitative methods and genome sequence	88
5.3.2 Quantification at three different levels (replcation domain/genes, chromosomes, genome)	88
5.3.2.1 Replication domain/gene level	89
5.3.2.2 Chromosome level	89
5.3.2.3 Genome level.....	89
5.3.3 Improvement of biological experiment design and image processing techniques	90
6. References	90
7. Abbreviations	110
8. Publication list	112
9. Acknowledgement	114

Chapter 1 Introduction

1.1 Translocation and spatial positioning

1.1.1 Cell nucleus and chromosomes

The mammalian cell nucleus is a membrane-bound organelle. It is bounded by a nuclear envelope, the inner and outer nuclear membranes are fused together to form nuclear pores in order to transit materials between the nucleus and cytoplasm (Stoffler et al. 1999). The nuclear pore complex has a basket that extends into the nucleoplasm. The peripheral nuclear lamina, which is composed of lamins A/C and B, lies inside the nuclear envelope and plays a role in regulating nuclear envelope structure.

Within the nucleoplasm, there are internal patches of lamin protein (Moir et al. 2000) and number of specialized domains or subnuclear organelles which are dynamic. There is rapid protein exchange among many of the domains and the nucleoplasm (Misteli 2001). CTs and subnuclear organelles have their characteristic 3D positions in nucleus (**Figure 1.1**).

Within the nucleus, the chromosomes are arranged into chromosome territories (CTs), with active genes on the surface or inside of the loosely packed territories (Cremer et al. 2000, Gilbert et al. 2004 (B)). Generally, chromosome homologues are not paired in interphase nuclei. In some cell types, heterochromatin is inactive chromatin anchored with the nuclear lamina or in internal nuclear regions. Above the level of the 30 nm fibre, eukaryotic chromatin is constrained into loops. Gene-rich domains are enriched in open chromatin fibers (Gilbert et al. 2004 (A)).

Chromosomes in a highly condensed state that normally occurs only for a brief period are always depicted as they are shown in **Figure 1.2**: in human somatic cells, chromosome 1 through chromosome 22 are numbered in approximate order of size. Two of each of these chromosomes plus two sex chromosomes --- two X chromosomes in a female; one X and one Y chromosome in a male are shown here. In this figure, the knobs on Chromosomes

13, 14, 15, 21 and 22 indicate the position of genes that code for the large ribosomal RNAs.

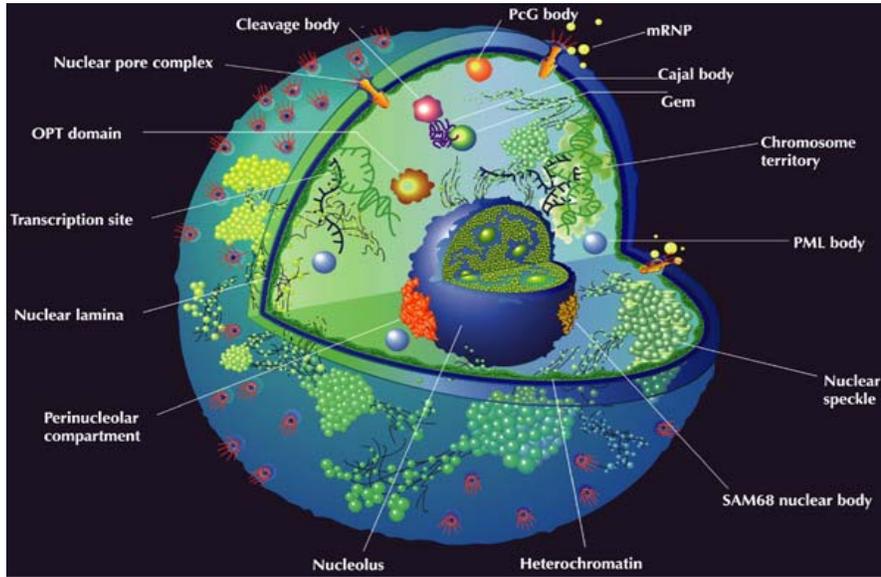


Figure 1.1 Cartoon of a mammalian cell nucleus showing the 3D position of chromosomes and a large number of nuclear domains (Spector 2001).



Figure 1.2 The banding pattern of human chromosomes stained with Giemsa at an early stage in mitosis. The regions of dark bands are rich in A-T nucleotide pairs. The

horizontal line represents the position of the centromere, which appears as a constriction on mitotic chromosomes. Adapted from (Franke 1981).

1.1.2 Cancer and translocation

In human cancers more than 600 recurrent balanced chromosomal rearrangements have been documented (Mitelman 2000). Chromosome abnormalities are described to be a cause rather than an effect of cancer (Rowely 2001). Specific recurring chromosomal translocations are often associated with a particular type of leukaemia, lymphoma or sarcoma (Rowely 2001).

Specific chromosomal translocations have consistently been associated with specific cancers (Elliott et al. 2002, Parada et al. 2002). For example, Chronic Myelogenous Leukaemia (CML) is a hematologic cancer which is caused by the $t(9;22)(q34;q11)$ translocation, producing the Philadelphia chromosome (Ph chromosome) and BCR/ABL fusion protein. Burkitt's Lymphoma (BL) is a hematologic cancer characterized by the $t(8;14)(q24;q32)$ in over 80% of patients; and Desmoplastic small round cell tumor (DSRCT) is an aggressive solid tumor associated with the recurrent $t(11;22)(p13;q12)$ translocation (Elliott et al. 2002).

The first genetically defined translocation is $t(8;14)(q24;q32)$, which involves the *MYC* and *IgH* genes on chromosomes 8 and 14, respectively (Zech et al. 1976, Dalla-Favera et al. 1982). Thereafter, many translocations were genetically identified and have been characterized by molecular methods (Rowely 2001). The cloning of translocation breakpoints has proved to be one of the most efficient ways to identify new genes that might be involved in cell growth regulation and malignant transformation inducement.

When DNA in one chromosome undergoes a double-strand break, the two ends at the break point may rejoin, or be attacked by nucleases, which leads to gene deletion or chromosome loss, or the ends may persist at least until the next cell division (Pederson 2003). When a second chromosomal break has occurred at the same time, the four ends on the two chromosomes may reseal into their original chromosomal axes, or both ends from one breakpoint join with those from the second chromosome, this last event constituting the cytogenetic event called reciprocal translocation (Pederson 2003).

In addition, chromosome translocations can also be detected from the cells of healthy individuals (Uckun et al. 1998). Although there is no follow-up studies of these individuals to determine if these cells became malignant, the translocation in these cells suggests that the presence of a translocation is not sufficient for a fully malignant phenotype.

1.1.3 Translocation and gene position

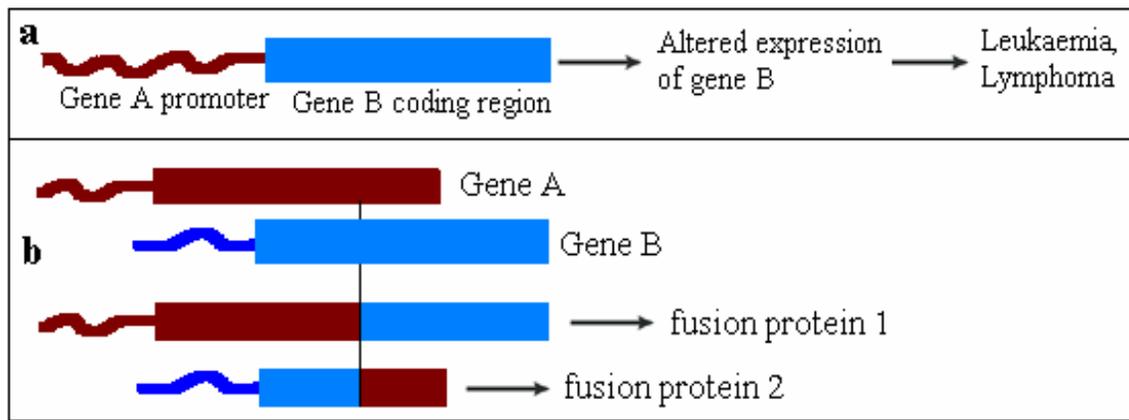


Figure 1.3 The juxtaposition and fusion pattern of genes in recurring chromosome translocations (changed from the figure in review Rowley 2001). (a) The juxtaposition of promoter/enhancer elements from one gene (gene A, red) with the intact coding region of another gene (gene B, blue). This happened in some lymphomas and leukaemias. (b) The recombination of the coding regions of two different genes results in one or two fusion proteins that might have new function. This happened in the translocations seen in CML and many of acute leukaemias, for example, the BCR–ABL fusion protein encoded by the Philadelphia chromosome.

The ends of one broken chromosome can join with the ends from the second broken chromosome. There are two possibilities for them to join together. As shown in **Figure 1.3** (a), malignant chromosome transformations can be induced by the juxtaposition of promoter/enhancer elements from one gene (gene A, red) with the intact coding region of another gene (gene B, blue) in some lymphomas and leukaemias. For example, the t(14;18) translocation, which is associated with follicular small cleaved-cell lymphoma,

involves immune receptor genes and juxtaposes the promoter region of *IGH* with the coding region of the anti-apoptotic protein *BCL2* on chromosome 18 (Fukuhara et al. 1979, Tsujimoto et al. 1984). Or, as shown in Figure 1.3 (b), two coding regions fuse to create a new, chimeric gene that encodes a fusion protein as seen in CML and many of the acute leukaemias and sarcomas (Shtivelman et al. 1985, Jaffe et al. 2001).

The spatial proximity of two chromosomal sites, or the two loci, can contribute significantly to their probability of undergoing translocations (Misteli 2004). If proximity affects the choice of translocation partners, the two loci and the two chromosomes involved in the translocation should be close to each other before the translocation (Parada et al. 2002).

The term 'position effect' was introduced by A. Sturtevant in 1925 for the phenomenon in which the action of a gene is altered by virtue of its location on a chromosome (Sturtevant 1925, Pederson 2003). Placing a gene next to heterochromatin may inactivate the gene, producing the effect known as position effect variegation, whereby a particular gene may be switched on in some cells and switched off in others. Such effects are turning out to be surprisingly widespread (Sumner 2003).

In last several years, the idea 'gene position effect' has rearisen. Individual gene loci have distinct local environments because of the presence of nonrandomly positioned chromatin domains and subnuclear compartments. The gene action may be influenced by its particular nucleus location (Parada et al. 2002, Shannon 2003). For example, in mouse lymphocytes, the dynamic repositioning of genes away from the chromocenter clusters or towards them is in association with gene activation or silencing (Brown et al. 1997, 1999). Giant chromatin loops which contain gene clusters were reported to extend outwards from corresponding chromosome territories surface according to gene(s) activation (Volpi et al. 2000; Williams et al. 2002, Mahy et al. 2002). By an analysis of the propinquity (or proximity) of selected gene loci in the nuclei of human lymphoblastoid cell line, Roix *et al.* reported that the closeness of certain pairs of gene loci in the nucleus is positively correlated with their propensity for oncogenic translocations (Roix et al. 2003).

The relative positions of BCR and ABL in haematopoietic cells are found in close proximity more often than by chance (Lukasova et al. 1997, Neves et al. 1999). Also,

RARA and PML are often close to each other. Both genes are frequently found to be fused in a t(15;17) translocation resulting in promyelocytic leukaemia (Neves et al., 1999).

The gene position relative to nuclear envelope and other various proteinaceous nuclear compartments (such as the nucleolus and a multitude of small nuclear bodies) is a fundamental property of every gene (Misteli 2004). Gene spatial positioning has functional roles both in gene activity and genome stability (Misteli 2004). Some nucleus location-dependent gene activities are believed (in many cases, assumed) to include replication, repair, transcription and breakage-rejoining events (Pederson 2003).

However, the functional significance of gene spatial positioning is poorly understood. Since it is believed that proximity is an important contributor to the possibility for a translocation to occur, it is helpful to do quantitative evaluation to analyze the proximity and the 3D distribution of genes in order to understand the translocation mechanism in a better way. For example, in spite of the considerable linear distance of ~30 Mb apart on chromosome 10 between gene *RET* and *H4*, which lead to radiation-induced thyroid tumors, the 3D localization analysis has demonstrated that these two loci are juxtaposed in the interphase nucleus of normal thyroid cells (Nikiforova et al. 2000).

In order to describe the position and even the spatial distribution pattern of genes accurately, it is necessary to develop some computational tools to quantify the 3D positions of the genes involved in translocation. The methods to measure distances among different genes, to perform a quantitative analysis of the positions of chromosome and genes, and to do statistical analysis have been developed by others (for example, Roix et al. 2003, Kozubek et al. 1999, Parada et al. 2002, Tanabe et al. 2002, Chambeyron et al. 2004, Bolzer et al. 2005). However, most analysis work published so far is concerning the 2D position analysis of chromosomes and genes in fixed nuclei. It is necessary to develop new quantitative computational tools to address the spatial distribution of genes involved in chromosome translocations in fixed nuclei and in live cells.

1.1.4 Gene position, proximity effects and chromosome position

The non-random positioning of chromosomes in the interphase nucleus has implications for the formation of chromosome translocations. For example, in order to undergo a reciprocal translocation, each of the two chromosomes must contain a free end generated by a double-strand break. The free chromosome ends can undergo limited diffusional motion (or some other motions) and become rapidly immobilized upon breakage (Lisby et al. 2003, Aten et al. 2004). A broken chromosome end can only have illegitimate joining with its immediate, nonrandomly positioned neighbors in order to give a reciprocal translocation (Elliott et al. 2002). Chromosome territories that are close to each other in the interphase nucleus have a higher chance of undergoing reciprocal translocations than chromosomes that are far apart (Parada et al. 2002).

The spatial proximity of two chromosomal sites, or the spatial proximity of two loci, can contribute significantly to their probability of undergoing translocations. The nonrandom physical position of the translocation partners contributes to the frequency of translocation as the genome regions frequently translocated in promyelocytic leukemia, acute myelocytic leukemia, Burkitt's lymphoma, and thyroid lymphoma were all found to be preferentially positioned in closer spatial proximity than nontranslocating regions in the same cell (Misteli 2004). Therefore, proximity effect is important to affect translocation frequencies and to determine translocation partners (Sachs et al. 1997).

On the other hand, by the analysis of chromosomal translocations, the arrangement of chromosomes in the nucleus can be investigated. This is one new approach emerged recently although it has not yet yielded a completely satisfactory picture (Perderson 2004, Cornforth et al. 2002). Typically, chromosome location in the nucleus is studied by Fluorescent in situ hybridization (FISH) with chromosome-specific probes. But the studies involved in this new approach have shown that the likelihood of a translocation is proportional to the distance between two chromosomal loci, the end-to-end ligation of fractured chromosome ends requires atomic juxtaposition (Kozubek et al. 1999, Roix et al. 2003). These studies have demonstrated how the probability of a reciprocal translocation relates to the intergenic distance of the two affected loci in chromosomes (Kozubek et al. 1999, Roix et al. 2003, Cornbluth et al. 2003).

1.1.5 Spatial positioning and clinical medicine

Clinicians found that the chromosome abnormalities were useful prognostic indicators (Mitelman 1981, Bloomfield et al. 1984). Until now the identification of chromosome aberrations is still the best known way to predict how a patient will progress or respond to treatment, so many clinics carry out the karyotyping on the cells of a leukaemia patient before treatment.

The difference between high versus low probability translocations involves distance in the range of 0.1 μ m. This positional relationship is functionally related to an Angstrom distance bimolecular collision-dependent covalent chemistry step. As has been pointed out (Pederson 2003), it is possible that this parameter (the distance in the range of 0.1 μ m), if expanded into a sufficient array of clinical material, might improve the detection of the pre-malignant state.

The Ph chromosome story is one successful story that was applied into clinical medicine. It began from the observation of a chromosome abnormality, the identification as a translocation, molecular analysis of the genes involved, to the functional characterization of the genes and finally to a drug that specifically targets the defective gene product (Rowley 2001). Undoubtedly new computational methods and tools about spatial positioning of chromosomes and genes will help clinicians to predict how a patient will progress or respond to treatment and apply this story into clinical medicine more accurately.

The techniques like FISH and spectral karyotyping (SKY) have greatly improved our ability to identify chromosomal abnormalities in cancer cells. New quantitative evaluation tools will provide us a new way to detect the pre-malignant state. Research of spatial position might have the potential to be applied for the dozens of translocations and chromosome abnormalities that are associated with human leukaemia, lymphoma, sarcomas and some benign tumors.

1.2 Chromosome positioning

1.2.1 Why is chromosome positioning important?

1.2.1.1 Chromosome positioning and gene expression

Chromosome positioning research is important because it may play important role in gene expression.

While a spatial organization of chromatin in the nucleus is well established, its functional significance is unclear (Chubb et al. 2003). However, it is possible that positioning of chromosomes can be a mechanism to increase gene expression efficiency and to process events (for example, gene expression) by creating specialized nuclear neighborhoods.

Some evidence showed that chromatin localization plays a role in the regulation of gene expression (Csink et al. 1996, Brown et al. 1997, Andrulis et al. 1998). Other evidence about nuclear architecture and the spatial organization of the genome showed that the higher-order chromatin structure and epigenetic regulation via chromatin modifications play important roles in gene expression (Francastel et al. 2000, Strahl et al. 2000, Dundr et al. 2001, Misteli 2001, Pederson 2001, Pederson 2002, O'Brien et al. 2000, Spector 2003). There is also evidence that the spatial positioning and nuclear architecture can help to establish the transcriptional states of telomeres in yeast (Feuerbach et al. 2002).

1.2.1.2 Gene position near heterochromatin or in nucleolus

Another reason is that, the position of a gene relative to other chromosomal loci within the nucleus can be important for its activation status. It has been shown, that the association of a gene next to heterochromatin is important for its silencing. For example, the β -globin locus is silenced when associated with heterochromatin, and several lymphoblast-specific genes are inactivated when they are near heterochromatin (Brown et al. 1997, Francastel et al. 1999). In yeast, the association of the loci with a peripheral region containing high local concentrations of silencing factors correlates with silencing (Maillet et al. 1996).

In nucleolus (and other subnuclear structures), there are gene clusters and chromosome clusters. Chromosome territories can be positioned to place specific genes in special gene

or chromosome neighbourhoods in order to favour the expression or silencing of genes. The neighbourhood of chromosomes or genes might provide a specific environment for the expression of some genes but not others, in order to exert regulatory functions (Lemon et al. 2002)

However, it is not clear how exactly heterochromatin contributes to gene silencing. We still do not know if heterochromatin exerts its inhibitory effect by blocking the access of transcriptional activators to the silenced domains. In yeast, silenced chromatin allows activator binding although silencers are required to maintain heterochromatin (cheng et al. 2000, Sekinger et al. 2001).

1.2.2 The history of chromosome positioning research

The first speculation about how the chromosomes are arranged in the nucleus was done at the end of the 19th century. In 1885, based on the studies of salamander cell division, Carl Rabl proposed that the centromere–telomere orientation of chromosomes observed during anaphase is maintained throughout the cell cycle (Rabl 1885).

Later, in 1909, based on his observations of blastomere nuclei in the nematode *Parascaris equorum*, Theodor Boveri developed several hypothesis, two of them are: (1) chromosomes occupy distinct chromosome territories in the cell nucleus, (2) CT order is stably maintained during interphase (Boveri 1909, Walter et al. 2003).

Throughout the interphase, each chromosome occupies a finite intranuclear volume from which all the other chromosomes are excluded. Every chromosome represents a structural unit referred to as a chromosome territory (CT). The non randomly distribution of chromosomes in the interphase nucleus reflects the distinct physical nature. A chromosome is not randomly distributed throughout the nucleus but occupies chromosome territory.

Thereafter, the morphological nature of chromosome positioning during interphase remained elusive for many years until the subject was addressed again by several pioneering studies in the modern era (Comings 1980, Manuelidis 1985). Using ultraviolet-laser microirradiation technology, Cremer and his colleagues confirmed the territorial organization of chromosomes in interphase cells elegantly in 1982 (Cremer et

al. 1982). The notion of CT was confirmed and extended by chromosome painting (Lichter et al. 1988); this technology enormously accelerated study of chromosome territories.

FISH studies showed more evidence for the existence of chromosomes as individual territories in higher organisms (Schardin et al. 1985, Pinkel et al. 1986). Chromosomes can easily be seen in cells not only during mitosis, when they appear as distinct, highly condensed entities, but also in interphase nuclei. High-resolution light microscopy and electron microscopy demonstrated that chromosome territories are distinct entities and there is no significant intermingling among them (Hochstrasser et al. 1986, Visser et al. 1999, Visser et al. 2000, Parada et al. 2002).

Chromosome territories can be visualized by in situ hybridization using fluorescently directly labelled probes specific for a single pair of chromosomes. Some pioneering methods were later introduced for visualizing chromosomes in living cells (Schermetelle et al. 2001). All of these advances made the study of chromosome spatial arrangement intensive during the last several years.

A significant step was made in 1999 when it was shown that two human autosomes, chromosomes 18 and 19, occupy relatively peripheral versus central locations, respectively, in the nuclei of human dermal fibroblasts (Croft et al. 1999). The location of human chromosome 18 in skin fibroblasts changed from a peripheral to a more central position when cells exited the cell cycle, and adopted its more peripheral location again when cells re-entered the cell cycle (Bridger et al. 2000) (But this conclusion was not confirmed by others (Bolzer et al. 2005)). Another study presented evidence that all of the more gene-rich chromosomes had a relatively central location and that the gene-poorer ones had more peripheral locations both in dermal fibroblasts and lymphoblastoid cell lines (Boyle et al. 2001). These studies contradict data about size-correlated distribution of chromosomes described for flat, adherently grown cells, such as fibroblasts, epithelial cells, etc. (Sun et al. 1999, Cremer et al. 2004, Bolzer et al. 2005).

Two widely-accepted conclusions were drawn from all these studies: One is that the interphase chromosomes occupy relatively defined locations in the nucleus; the other is that the interphase chromosomes display a degree of translational freedom, moving within a confinement volume (Pederson 2004).

The next issue for the research of CTs is how territorial arrangement ordered during different cell divisions and among different cell types is ordered.

1.2.3 The tethering of interphase chromosomes

Some chromosome subregions show extensive mobility in nucleus, and the rapid movements of chromosome loci were reported in nuclei of budding yeast and of *Drosophila* embryos (Csink et al. 1998, Marshall 2002, Gasser 2002). However, interphase chromatin demonstrates relatively immobile by some sort of nuclear tethering (Abney et al.1997, Bornfleth et al.1999, Manders et al. 1999, Dundr et al. 2001, Gerlich et al. 2003, Walter et al. 2003). Live cell imaging of fluorescently labelled chromosomes indicates that chromosome territories undergo very limited large-scale translational motion during late G1, S and G2 phases (Abney et al. 1997, Zink et al. 1998, Manders et al. 1999, Vazquez et al. 2001).

Chromatin located at the periphery or nucleoli is significantly less mobile than that at the nuclear interior (Chubb et al. 2002). Although it is possible that the large excluded volume taken up by all chromosomes relative to the limited total volume of the nucleus contributes to the immobility of chromosomes, there are lots of additional physical constraints which prevent chromosomes from moving and therefore contribute to the maintenance of chromosome positioning in the interphase nucleus (Parada et al. 2002).

There are multiple interaction sites between chromosomes and relatively immobile features of nuclear architecture. For example, as shown in **Figure 1.4**, it is possible that chromosome can be tethered with nuclear envelope, Nuclear Pore Complex (in yeast but not in mammalian cells), lamina, nuclear matrix-attachment regions, subnuclear compartments and other nuclear structures like pre-mRNA splicing-factor compartments (Parada et al. 2002).

a. Nuclear envelope (NE) and lamina

NE is the most obvious substrate to tether chromatins. Studies in yeast and mammals provided evidence that the association of chromatin with NE constrains its translational motion (Heun et al. 2001, Chubb et al. 2002). There are also convincing evidence to show that chromosomes attach to the nuclear lamina (Goldman et al. 2002). The observation

that the major component of the intermediate-filament-like lamins can bind directly to DNA and core histones, provides evidence that chromatin can be tethered to the lamina (Luderus et al. 1994, Taniura et al. 1995, Goldberg et al. 1999). Moreover, study showed that it is also possible for chromosomes to interact with and to be tethered by internal lamins (Goldman et al. 2002). However, a minority opinion is that these attachments involve some ill-defined internal nuclear structure (Pederson 2004).

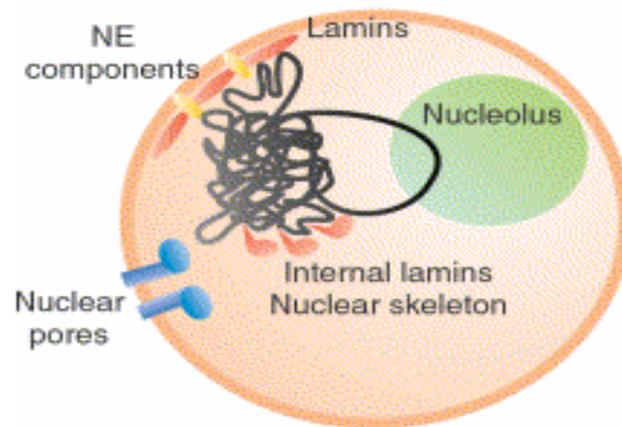


Figure 1.4 The tethering of interphase chromosomes (Parada et al. 2002 (B)).

b. Nuclear Pore Complex (NPC)

In yeast, telomeric chromatin regions are tethered to the nuclear pores via the Mlp2 protein and there is specific anchorage of telomeres at the nuclear periphery (Andrulis et al. 1998, Galy et al. 2000, Feuerbach et al. 2002). There is also evidence demonstrating that the yeast telomere protein yKu70 mediates the tethering of telomeric regions to nuclear pores, and pore tethering via Nup2 acts as a boundary element and blocks the spreading of heterochromatin (Laroche et al. 1998, Ishii et al. 2002). But for mammalian cells, there is still no evidence for the direct physical interaction of chromatin with nuclear pores.

c. Nuclear matrix-attachment regions

It is also possible for chromosome to be restrained by internal nuclear structures. Along the entire length of chromosomes, nuclear matrix-attachment regions have been identified

, but the protein-binding partners of matrix-attachment regions keep elusive (Gasser et al. 1986).

d. Nucleolus

The acrocentric chromosomes (13-15, 21 and 22) which have tandemly repeated ribosomal genes (rDNA) known as the nucleolar organizing region (NOR) cluster together in the nuclear space to form the nucleolus around the highly expressed rDNA. The chromosomes that contain NOR and rDNA are restrained from movement by their association with the nucleoli (Chubb et al. 2002).

e. Other nuclear structures

Other nuclear structures (pre-mRNA splicing-factors, nuclear proteins) might also have restraining forces for chromosomes. The positions of pre-mRNA splicing-factor compartments are stable, so they can act as anchoring points or at least as 'roadblocks' for the movement of chromosome territories (Eils et al. 2000). Nuclear proteins can also play a role to restrain the movement of chromosome: when mammalian cell nuclei are digested with restriction endonucleases, some of the DNA sites that should be susceptible to this enzyme keep uncleaved, demonstrating that there is one kind of tight binding of nuclear proteins to these DNA sequences.

In general, it is believed that the tethering of chromosome territories to various nuclear structures contributes to the maintenance of chromosome positioning, thus restricts their movement. Understanding these tethering sites will be helpful for the quantitative analysis of gene/chromosome positioning during interphase, mitosis, and different cell divisions.

1.2.4 Why is chromosome positioning research difficult?

Powerful technologies like FISH and SKY can help us to visualize gene loci in any chromosomes. The advances in genome sequencing, light and electron microscopy allow us to detect almost any details of one chromosome. However, it is still difficult to answer how chromosomes are arranged in the nucleus. There are several reasons:

Firstly, different cell types can vary widely and chromosomes are arranged within the nucleus in different way. It is impossible or at least difficult to find a universally valid

principle for the chromosome arrangement during interphase. For example, Rabl configuration is the characteristic of several plant cell types, but not typical for most mammalian cells (von Driel et al. 2004).

Secondly, within each cell type, changes in gene activity, differentiation stage and cell cycle status can also result in alterations of nuclear architecture (Haaf et al. 1991, Brown et al. 2001, Bridger et al. 2000).

Finally, the three-dimensional nature of interphase nuclei, the relatively large number of chromosomes and the absence of fixed reference points within the nucleus make spatial pattern recognition and analysis difficult (Parada et al. 2002).

Collectively, these differences and difficulties suggest that we must be extremely cautious in getting any conclusion of one cell type from any other particular cell type. As there are a few general features of nuclear organization that appear to be relevant to understand regulated gene expression, new computational tools should be developed to detect and explain the general features and rules behind the different experimental systems.

1.3 Image processing, computer vision and the applications

1.3.1 Image processing and computer vision

Vision allows us to see and understand the real world. However, (biological) image data captured by artificial sensors (such as CCD camera) are generally complex, full of noise and not easy to interpret by pure visual inspection. To improve image data and to be able to interpret them, it is necessary to apply image processing methods for object detection, motion estimation, visualization, and quantification. Although image processing does not increase the image information content, it helps us to understand the image data and to quantify the information in the images in a better way.

The following sequence of image processing steps is commonly seen in quantitative biological statistics (Sumner 2003, Gerlich et al. 2003):

- a. An image is captured by a sensor and digitized. Raw images are disturbed by noise caused e.g. by specimen preparation (the unspecific binding of a probe, or the autofluorescence of the specimen), illumination conditions, camera settings, display parameters, or the conversion of light signal into electrical signal in the CCD camera or other recording device.
- b. In order to get rid of the different types of noise, especially the high-frequency noise which alters the image contours and causes the loss of sharpness, several kinds of filters can be used for further image processing: linear convolution filters like low-pass filter, nonlinear filters like median filter, or more sophisticated filters like anisotropic diffusion filter. Image preprocessing methods (different filters) are applied to try to enhance some object features, which are relevant to understanding the image in a better way.
- c. Image segmentation is the next step to separate objects from the image background and from each other. Automatic segmentation is necessary for the large amount of 4D (three-dimensional images over time) or 5D (4D plus channels) live cell image data produced by confocal microscope. There are two different approaches for segmentation: Contour-oriented segmentation and region-oriented segmentation. Based on the image gradients, contour-oriented segmentation detects the differences between neighboring image points. Region-oriented segmentation determines the similarities of neighboring pixels that are merged and assigned to image segments. However, as both approaches strongly depend on the specific application, the performance of both approaches is often very different, which can lead to different results.
- d. Other methods of image processing, like object description, classification and tracking in a totally segmented image, are also necessary for the analysis of data. Extracting information from moving scenes are important technologies in image processing, but they are not discussed in this thesis.

Computer vision aims to duplicate the effect of human vision by electronically perceiving and understanding an image. The discipline of computer vision is growing in both theory and applications during the last forty years: the research articles have grown beyond possible prediction, and there are more and more applications that bring vision into

contact with the wider public: teleconferencing, video phones, and the use of imaging of many modalities in medical diagnosis (Sonka et al. 1998).

In order to simplify the task of computer vision understanding, two levels are usually distinguished: low-level image processing and high-level image understanding. Low-level methods often include image compression, pre-processing methods for noise filtering, edge extraction, and image sharpening. Low-level computer vision techniques overlap almost completely with digital image processing discussed above, which has been practiced for decades. High-level processing is based on knowledge and goals, and artificial intelligence methods can be applied. Computer vision is expected to solve very complex tasks in biological systems.

1.3.2 Image processing applications in biological data

Biological 3D or 4D image data are generally complex and can consist of thousands of individual image slices, which can occupy a lot of gigabytes of storage. The development and the application of computer vision methods and other computational tools allow researchers today to detect the dynamics of complex structure in live cells and organisms. The quantitative analysis and visualization can be performed in a large scale, and based on these results mathematical models can be developed which are helpful to verify the conclusions about the mechanism behind the observed biological phenomena.

1.3.2.1 Analyzing the dynamics of cell image data

All biological phenomena are dynamic. Live cell studies have revealed that there are high dynamics of many nuclear subcompartments that were previously thought to be stable (e.g., Misteli et al. 1997, Marshall et al. 1997, Platani et al. 2000).

A simple approach to estimate the dynamic, morphological alterations of live cell data over time or the velocity with which an organelle moves within the cell is to interpret time-lapse movies by visual inspection. This is helpful to obtain an overall or global impression of motion patterns, however, it is not suitable to quantify the data in cells and to address underlying mechanisms.

Therefore, a convincing presentation of functional assays addressing cell nuclear structures and dynamics under different experimental conditions requires quantitation by automated image processing techniques. There are already some successful cases, in which the dynamics of several nuclear subcompartments were observed to be regulated within the cell (Misteli et al. 1997, Eils et al. 2000, Muratani et al. 2002, Vazquez et al. 2001, Heun et al. 2001, Platani et al. 2002).

1.3.2.2 Quantitative analysis and visualization of biological image data

The techniques for quantitative analysis and visualization of live cell imaging data were developed during the last several years. By the dynamic analysis of several nuclear subcompartments, their applicability was exemplified (Gerlich et al. 2003).

Quantitative tools for the measurements of lots of parameters directly from the images (or image stacks) are now available. According to the gray values in the segmented area of corresponding original images, we can determine the amount and concentration of fluorescently labeled protein; then the velocity of the detected object or each point on the object surface or in space can be determined. Thereafter, parameters like acceleration, bending or tension, volume changes, concentrations, or diffusion coefficients can be determined, and statistical analysis can provide further possibilities for quantification.

1.3.2.3 Development of mathematical model

Mathematical models are important to test or verify a hypothesis and to develop further experimental designs (Phair et al. 2001). The mathematical description of different biological systems (e.g. mitosis, cell divisions, signal pathway) is important as it provides one way to integrate the results from many individual experiments into a complex system. The strategies in systems biology will identify some novel principles of cellular regulation and gene expression during interphase and mitosis from the huge amount of experimental data that were generated.

Data generated from live cell microscopy can be used to develop mathematical models that describe different biological processes. The quantitative methods presented in this thesis are a possibility to provide one building block for the description of the spatial distribution of genes and chromosomes. A challenge for future work of models is to

extract more parameters from the data. This has been exploited in medical image analysis where the human brain was modeled using finite element method, in order to understand the forces occurring during brain deformations (Ferrant et al. 2001).

1.3.3 How to apply image processing tools into chromosome/gene position data?

There are already some applications for the quantitative image analysis. For example, some studies showed that chromatin underlies slow diffusional motion, and this movement is confined to relatively small regions in the nucleus (Marshall et al. 1997, Vazquez et al. 2001, Heun et al. 2001, Bornfleth et al. 1999). Other studies showed that quantitative image analysis is helpful to provide evidence that, for PML bodies, Cajal bodies and chromatins, transport dependent on metabolic energy can occur in the nucleus (Muratani et al. 2002, Heun et al. 2001, Platani et al. 2002).

At the same time, a number of commercial software is available, which allow 3D visualization, segmentation, tracking and other image analysis to be done automatically.

For example, as stand-alone software, Amira (www.tgs.com) has a variety of image processing functions (image loading for microscopic data, image preprocessing, surface reconstruction, segmentation, 3D visualization, etc.). It is a powerful 3D visualization toolbox including volume rendering and graphical surface rendering techniques. A macro scripting language is included and basic tools for quantitative analysis are available.

Open Inventor (www.tgs.com), which owns a class library (C++ or Java) for image analysis, can realize the 3D visualization based on OpenGL with a perfect user interface. It is Operating System-independent and can be used to visualize and analyze the 3D or 4D (3D + time) complex datasets.

ImageJ (<http://rsb.info.nih.gov/ij/>), can do image preprocessing and processing for 2D + time data with limited visualization. As a free software, it is written in Java, user can develop applets or applications by themselves. Many plugins for ImageJ are available from the Internet.

TILL visTrack (www.till-photonics.de), which comes as a plugin for Olympus microscope systems, can do quantitative analysis for 2D + time data set. It is only available for the Windows operating system.

Bitplane Imaris (www.bitplane.com), which can do 3D or 3D + t visualization and quantification, is a stand-alone and commercial software with powerful image analysis tools and limited tracking function.

The software mentioned above have a lot of advantages and can help people to analyze huge amount of data. However, they do not provide enough options to allow the researchers to address specific questions in some biological data sets, especially for chromatin spatial distribution data sets.

Based on the studies and software mentioned above, several steps in image processing are used in this thesis:

a. Preprocessing

Images acquired by confocal microscopy are full of noise. It is necessary to apply some preprocessing steps in order to get the images with low signal-to-noise from raw images, especially with low high-frequency noise which causes loss of sharpness and alters the image contours.

b. Segmentation and surface reconstruction

Segmentation subdivides an image into its constituent parts (in the case of our data, chromosomes or gene loci), which are defined as homogenous and disjoint regions (image segments) that are separated by boundaries. In order to quantify the spatial distribution of chromosomes or genes, it is necessary to extract the surface of these entities from image stacks (Surface reconstruction could also be done based on voxels).

c. Quantitative analysis

Based on the steps described above, some quantitative analysis, for example object classification, tracking and parameter estimation, can be applied into chromosome/gene position data sets.

1.3.4 What is new in this thesis

During the last ten years, using fluorescent proteins (GFP and its variants) as a universal fluorescent marker to visualize many cellular structure virtually in living cells and fixed cells (Chalfie et al. 1994, Heun et al. 2001), the cell imaging technology and Fluorescence In Situ Hybridization (FISH)/multi-FISH have revolutionized the study of

cellular and nuclear structure and dynamics. New microscopes and related software are making it possible to rapidly record 4D images and even 5D images.

At the same time, based on methods from image processing, computer vision and other related fields, a lot of new approaches have been developed to evaluate the cell nuclear dynamics and the relative position of the genes, chromatin, replication loci and chromosomes. Image analysis has become a key tool for understanding the complex organization of biological processes in live specimens (Gerlich et al. 2003 (B)).

For example, sophisticated computational algorithms were developed to accurately compute chromosome volumes within the nuclear space (Eils et al. 1996, Visser et al. 1998). Based on cluster analysis in color space, Saracoglu et al. developed an approach for the automated analysis of M-FISH analysis, which allows the 3D analysis of M-FISH labeled chromosomes in interphase nuclei (Eils et al. 1998, Bolzer et al. 2005). Using 5-shell method (the area of cell nucleus was divided into concentric shells (1-5) of equal area from the periphery of the nucleus to the center), Croft et al. draw the conclusion that human chromosome 18 adopts a more extensive position associated with nuclear matrix, but chromosome 19 occupies a more internal position in the nucleus (Croft et al. 1999).

Gerlich and his colleagues (Gerlich et al. 2003 (B)) discussed the concepts for automated analysis of multidimensional image data from live cell microscopy and their application to the dynamics of cell nuclear subcompartments. Using contingency table λ^2 analysis and other statistical analysis, Roix et al. demonstrated that higher-order spatial genome organization is a contributing factor in the formation of recurrent translocations (Roix et al. 2003). Kreth et al. applied statistical methods and computer simulated arrangements of CTs to calculate interchange frequencies between all heterologous CT pairs, assuming a uniform action of the molecular repair machinery (Kreth et al. 2004).

However, these tools are always specific to (or concentrate on) only one cell type or one kind of experiments. People still lack of tools which are powerful enough to do more quantitative analysis for difference cell types. Although these methods help to understand the mechanism of nuclear dynamics in a better way, we are still at a highly speculative stage to address the underlying molecular mechanisms of cell nuclear process and chromosome abnormalities.

In order to do quantitative analysis of chromosome position and nuclear structure, it is necessary to develop more computational tools to analyze 3D (also 4D or 5D) images, as these new computational methods have become a key tool to understand the complex organization of biological processes in live specimens (Gerlich et al. 2003 (B)).

In this thesis, we present some new three-dimensional quantitative tools using the concept of object-oriented programming (<http://Java.sun.com/docs/books/tutorial/Java/concepts/>) and Java programming language, to describe, analyze and visualize the 3D distribution and the spatial arrangement of chromosomes and genes in fixed nuclei. At present, the prerequisite to use these new tools is that a 3D surface reconstruction of chromatin signals has to be obtained at first from confocal image stacks.

As one example to show how to do some quantitative analysis for these data sets, a method for deformable surface registration using thin-plate splines and some computational tools to calculate the distance and angles are presented in chapter 2 of this thesis, thereafter, two applications-oriented projects are presented in chapter 3 and chapter 4, respectively. From these two applications the biological conclusions were drawn, respectively. For the first time, the methods presented in this thesis provide one example in which point set registration allows the analysis and evaluation of chromosome spatial distribution inheritance during several cell divisions. And it is the first time that the spatial distribution of genes, with focus on gene *MLL* and some of its translocation partners, was analysed and quantified within the 3D space of interphase nuclei of hematopoietic cells.

Chapter 2 Computational methods for quantitative analysis and visualization of nuclear structure in fixed cell nuclei

2.1 The image analysis pipeline used in this work

In order to motivate the computational methods presented in this chapter, a general outline of the image analysis pipeline used in this work – from the acquisition of the data up to the biological conclusions – is presented first. Several of the following steps have been carried out by the biological collaboration partners (a, b, c, and partially d and f) whereas in this work the step e (and partially d and f) was focused which will be detailed in the remaining sections of this chapter.

- a. The raw biological image data that occupied hundreds of gigabytes of storage were produced by confocal microscope, to record the gene signals involved in different translocations, or the signals corresponding to chromosome territories in several human cell lines.
- b. For image preprocessing, the Gaussian filter or the Median filter was applied to all images to reduce noise.
- c. Threshold-based segmentation was applied where the threshold was selected manually by the biologists. Current state-of-the-art image segmentation software as far as tested in our framework failed to automatically segment our specific images in a completely correct way.
- d. Surface reconstruction was done semi-automatically by Amira 3.0, a software produced by IVC GmbH. These surface reconstructions were checked and revised by biologists and by the program developed by ourselves, see section 2.2 – 2.5 for details. Based on these surface reconstructions, methods from image processing and computational geometry were explored, evaluated and applied for these data. 3D visualization and statistical analysis have been applied thereafter. For details of these methods see the remaining sections of chapter 2 and section 3.3.

- e. In particular, a shape similarity measure, called Bending Energy has been used to evaluate the chromosome spatial distribution and chromosome position inheritance during cell divisions.
- f. Some biological conclusions were drawn from the quantitative analysis and the related models of chromosome positioning during interphase and mitosis were discussed, in order to understand the mechanism of translocation and chromosome position inheritance during cell divisions in a better way. For details of the discussion see chapter 5.
- Figure 2.1** shows the workflow of methods described in this chapter.

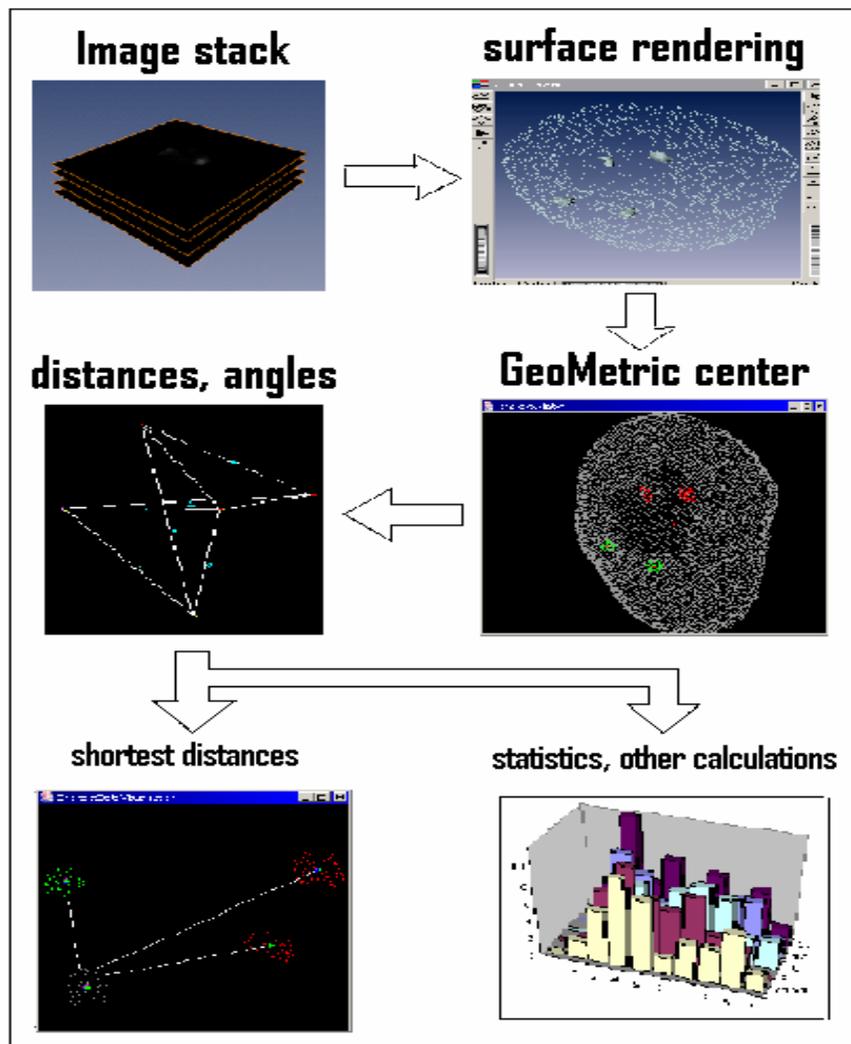


Figure 2.1 Work flow of methods described in this chapter.

2.2 Landmark data

When comparing shape a common approach is to describe the shape by points which are characteristic for it in the sense that they can be localized also on other objects (spatial entities) with the same kind of shape, for instance, other objects within the same population. In section 2.2.2, we will call these points landmarks. After, based on the pair-wise correspondence of the landmarks obtained from the shapes of two objects a measure quantifying the similarity of the shapes is defined. However, for comparing the arrangement of chromosomes in different nuclei it is not desired, as explained in the following section, to take into account the shape of each single chromosome territory. Hence, these arrangements are represented by the geometric center of each chromosome territory as defined in the following section.

2.2.1 Geometric center of spatial entity

In this thesis, spatial entities can be Chromosome Territories (CT), replication loci or genes. The geometric center (GC) of the surface points of one spatial entity is defined as:

$$\left(\frac{1}{n} \sum_{i=1}^n x_i, \frac{1}{n} \sum_{i=1}^n y_i, \frac{1}{n} \sum_{i=1}^n z_i \right).$$

Here, n is the total number of points extracted from the considered spatial entity, and x_i , y_i , and z_i are the coordinates of the points extracted from it, respectively. For the visualization of the geometric center see **Figure 2.2 (a)**.

In order to focus on the topologic arrangement of chromosome territories and to neglect the effects coming from differences in their shape, the geometric centers of these spatial entities are used instead of their surface points. This leads to notably different results as the shape of replication loci or genes, and especially of homologue CTs can significantly vary in different nuclei even when the relative positions of their geometric centers are identical.

All calculations presented in this thesis are based on the geometric center of spatial entities, therefore, the mathematical analysis involved ignores the spatial extensions of the considered objects. However, the shape of replication loci or genes, and especially of

homologue CTs can significantly vary in different nuclei even when the relative positions of their geometric centers are identical. When the number of geometric centers describing the arrangement is too small, this can lead to a confusion of the quantitative measure (see section 5.1).

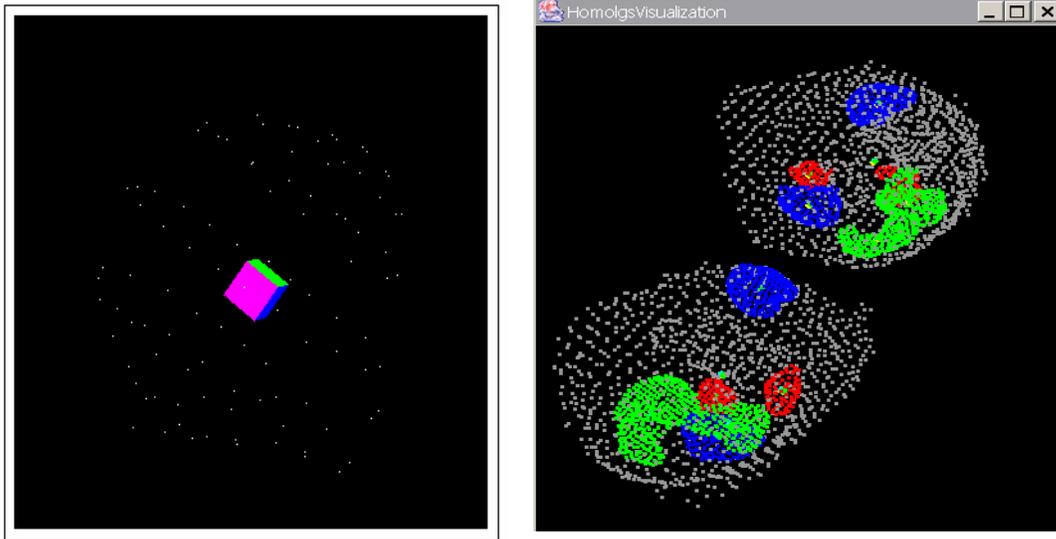


Figure 2.2 Visualization of geometric centers (a) and spatial entities (b) in Java 3D viewer. (a) The geometric center (the pink cube) of one homologue (white point set). (b) The visualization of several spatial entities (here in the case of homologue CTs) in two cell nuclei. White: point sets for human cell nuclei, green: chromosome 4, blue: chromosome 7, red: chromosome 21.

2.2.2 Landmarks - definition

A landmark is a point of correspondence on each object that matches between and within populations (Dryden et al. 1998). According to Dryden et al., there are three basic types of landmarks: anatomical points that correspond between organisms in some biologically meaningful way, mathematical landmarks that satisfy some mathematical extremal property, e.g., point of maximum curvature, and pseudo-landmarks constructed on an organism for which correspondence can be established algorithmically for different spatial entities.

A label is a name or number associated with a landmark, and identifies which pairs of landmarks correspond when comparing two objects. Such landmarks are called labelled landmarks. The landmark with label 1 on one specimen corresponds in some meaningful way with landmark 1 on another specimen. A labelling is either naturally apparent or an objective method of relabelling could be used in certain situations (Dryden et al. 1998).

2.2.3 Configuration space

The configuration is the set of landmarks on a particular object. The configuration matrix X is the $k \times m$ matrix of Cartesian coordinates of the k landmarks in m dimensions. The configuration space is the space of all possible landmark coordinates. In every data set of the applications involved in this thesis, there are $k > 5$ landmarks in $m = 3$ dimensions.

2.3 Analysis of landmark data by the means of angles and distances

The methods for quantitative analysis of chromosomes and genes (the distance and angle calculation) have been developed by others (e.g. Kozubek et al. 1999, Parada et al. 2002, Tanabe et al. 2002, Chambeyron et al. 2004). For example, in a two-dimensional analysis of human fibroblast prometaphase rosettes, Nagele et al. (Nagele et al. 1995, Nagele et al. 1998) measured distances and angular separations for a number of chromosomes. Angular measurements were made by Koss (Koss 1998) between the center of nucleus and homologous pairs of several CTs. He reported that in about two-thirds of the bronchial epithelial cell nuclei, the two homologous formed angles nearly identical to those reported by Nagele et al. (Nagele et al. 1995) for the same chromosome pairs in fibroblast prometaphase rosettes. However, the calculations in these studies focus on the distance and angle calculation in a two-dimensional analysis. It is necessary to develop new computational tools for the quantitative evaluation of chromatin spatial distribution based on 3D landmark data.

2.3.1 Size measure

The Euclidean distance between two landmarks (or geometric centers) is defined as:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

Here (x_1, y_1, z_1) and (x_2, y_2, z_2) are the 3D coordinates of two landmarks, respectively.

Consider X to be a $k \times m$ matrix of Cartesian coordinates of the k landmarks in m dimensions, i.e. the configuration matrix of the object, then a size measure $S(X)$ is defined as any positive real valued function of the configuration matrix such that

$$S(aX) = a S(X)$$

for any positive scalar a .

A Norm is a special size measure S which satisfies additionally among others: $S(X + Y) \leq S(X) + S(Y)$. There are several examples for size measure (see Dryden et al. 1998 for details):

(1) centroid size

The centroid size $S(X)$ is the square root of the sum of squared Euclidean distances from each landmark to the centroid, namely

$$S(X) = \sqrt{\sum_{i=1}^k \|(X)_i - \bar{X}\|^2}$$

where $(X)_i$ is the i th row of X ($i = 1, \dots, k$) and $\bar{X} = (\bar{X}_1, \dots, \bar{X}_m)$ is the centroid. This measure of size will be used throughout this thesis. Centroid size is the most commonly used size measure in geometrical shape analysis (e.g. Bookstein 1986, Kendall 1984, Goodall 1991, Dryden et al. 1992).

(2) maximum distance size S_{max}

S_{max} is defined as the maximum of the $k(k-1)/2$ inter-landmark distances.

(3) sum of inter-landmark distances.

2.3.2 The Hausdorff distance

As there is no common reference point for all spatial entities produced by the signals in confocal image stacks, it is difficult to describe the exact 3D position of one spatial entity. In order to describe the position \mathbf{x} of one spatial entity inside the nucleus, the "

Hausdorff distance" $d(\mathbf{x}, S_R)$ is used to get an indication about the position relationship between this spatial entity and nuclear surface S_R :

$$d(\mathbf{x}, S_R) = \min_{r \in S_R} |\mathbf{x} - r|$$

r is one point which belongs to the point set S_R . Hence, $d(\mathbf{x}, S_R)$ is the shortest Euclidean distance from the geometric center of spatial entity \mathbf{x} to the point set S_R , which is in the calculations used in this work a finite set of extracted points on the surface of the cell nucleus.

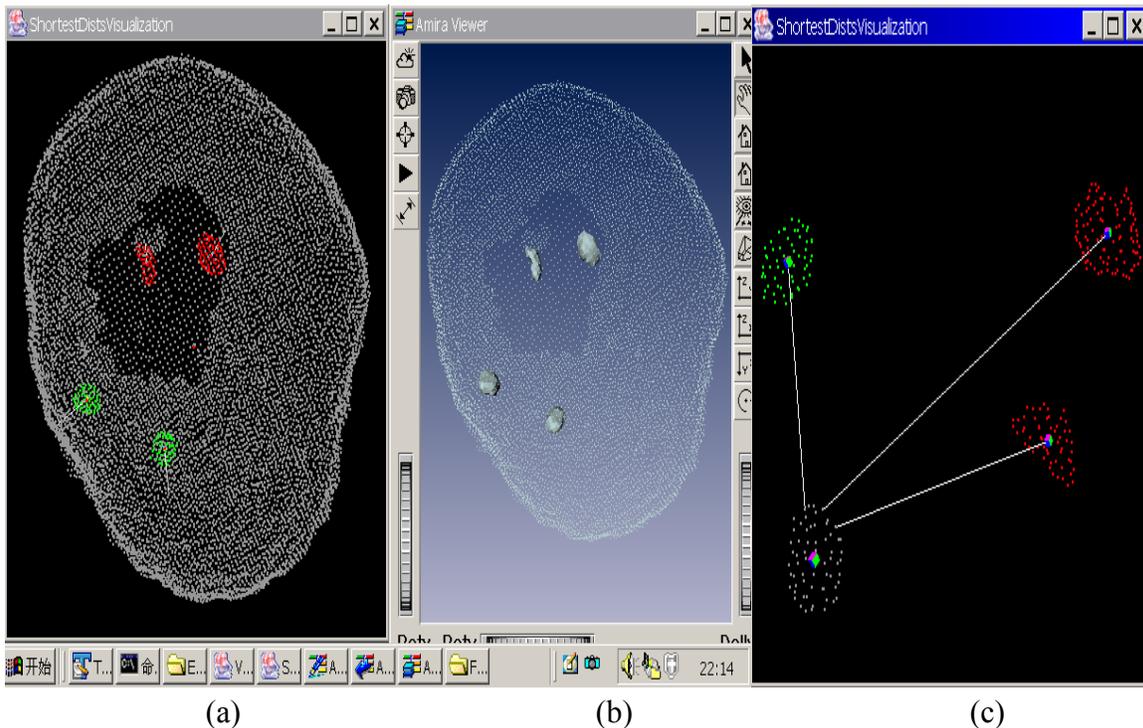


Figure 2.3 3D visualization of Hausdorff distance. (a) Hausdorff distance (*short white line*) of gene *MLL* (*red*) or *AF9* (*green*). Nuclear surface points are in white. (b) 3D surface reconstruction of gene *MLL* and *AF9* obtained directly from confocal image stacks using Amira 3.0. (c) Another example to demonstrate the Hausdorff distance between the geometric centers of spatial entity A (*red*), B (*red*) or C (*green*) and the surface points of spatial entity D (*white*).

By this way every Euclidean distance between the gene position \mathbf{x} and every nuclear surface point are calculated first, and then the smallest distance among these distances is used as the Hausdorff distance. In **Figure 2.3** (a), the Hausdorff distances are shown as

short white lines, which are calculated directly from the 3D surface reconstruction in Figure 2.3 (b). Figure 2.3 (c) shows another example in detail to calculate the Hausdorff distances between the geometric centers of red (or green) entities and the surface points of white spatial entity. In (c), the different Hausdorff distances for different entities (here in the case of genes) correspond to different surface points on the spatial entity in white.

2.3.3 Normalization of the configuration

Shape and size of cell nucleus are subject of permanent change and depend especially on the actual phase of the cell in the cell cycle but they can also vary considerably in different cells observed at the same phase in their respective cycle. As in this work the topological arrangement of the chromosomes is of interest rather than their relative geometrical positions we have to get rid of the variation in position due to global changes in shape and size.

In terms of a configuration, in the configuration space, a step towards this direction is to normalize this configuration according to a given size measure or a given norm, i.e. to multiply the configuration X with a scalar a such that the considered size S (or norm; see section 2.2.4) of aX equals to one ($S(aX) = 1$; as $S(aX) = aS(X)$ this implies $a = 1/S(X)$).

As we will see later on, if we compare two configurations using a bending energy based measure, we will get implicitly rid of global nuclear shape and size variations and no normalization is required. However, distance based methods to compare to configurations depend heavily on their size. Therefore we tried different size measures for the normalization in this thesis:

(1) Maximum Normalization (normalization with the size measure S_{max})

It corresponds to the normalization with the maximum norm in \mathbb{R}^1 . After normalization one normalized distance will be 1, which is the maximum distance of this configuration.

(2) Mean Normalization

Mean Normalization corresponds to the normalization with the mean norm in \mathbb{R}^1 . After normalization one normalized distance will be 1, which is the arithmetic mean distance of this configuration. In other words, given one group of data (the data can be distance, DD or DM) in one cell, the arithmetic mean value is calculated from this group, and every value of this group is divided by the mean value.

2.3.4 Distance Difference (DD) of two homologues in the same chromosome of two cells

Distance difference (DD) is defined as the difference of the distances of two homologues in the same chromosome of two cells. As shown in **Figure 2.4 (a)**, in the case of two homologues a and b in every chromosome, every homologue is represented as one point using its geometric center. In nucleus one, the distance between two homologues is 4; and in nucleus two, the distance between the two homologues of the same chromosome is 7; so the distance difference of this chromosome in two cells is $(AB - CD) = 7 - 4 = 3$.

In the case of three homologues in chromosome 7 of HeLa cell nucleus, as shown in **Figure 2.4 (b)**, the distance difference between the same chromosome in two cells is defined as the minimal value of six DD combinations (**Figure 2.4(c)**):

$$\text{Combination 1} = \text{Diff}(A1,A2) + \text{Diff}(B1,B2) + \text{Diff}(C1,C2);$$

$$\text{Combination 2} = \text{Diff}(A1,A2) + \text{Diff}(B1,C2) + \text{Diff}(C1,B2);$$

$$\text{Combination 3} = \text{Diff}(A1,B2) + \text{Diff}(B1,A2) + \text{Diff}(C1,C2);$$

$$\text{Combination 4} = \text{Diff}(A1,B2) + \text{Diff}(B1,C2) + \text{Diff}(C1,A2);$$

$$\text{Combination 5} = \text{Diff}(A1,C2) + \text{Diff}(B1,A2) + \text{Diff}(C1,B2);$$

$$\text{Combination 6} = \text{Diff}(A1,C2) + \text{Diff}(B1,B2) + \text{Diff}(C1,A2);$$

Here $\text{Diff}(X, Y)$ is the distance difference between homologue X and Y of the same chromosome in two different cells.

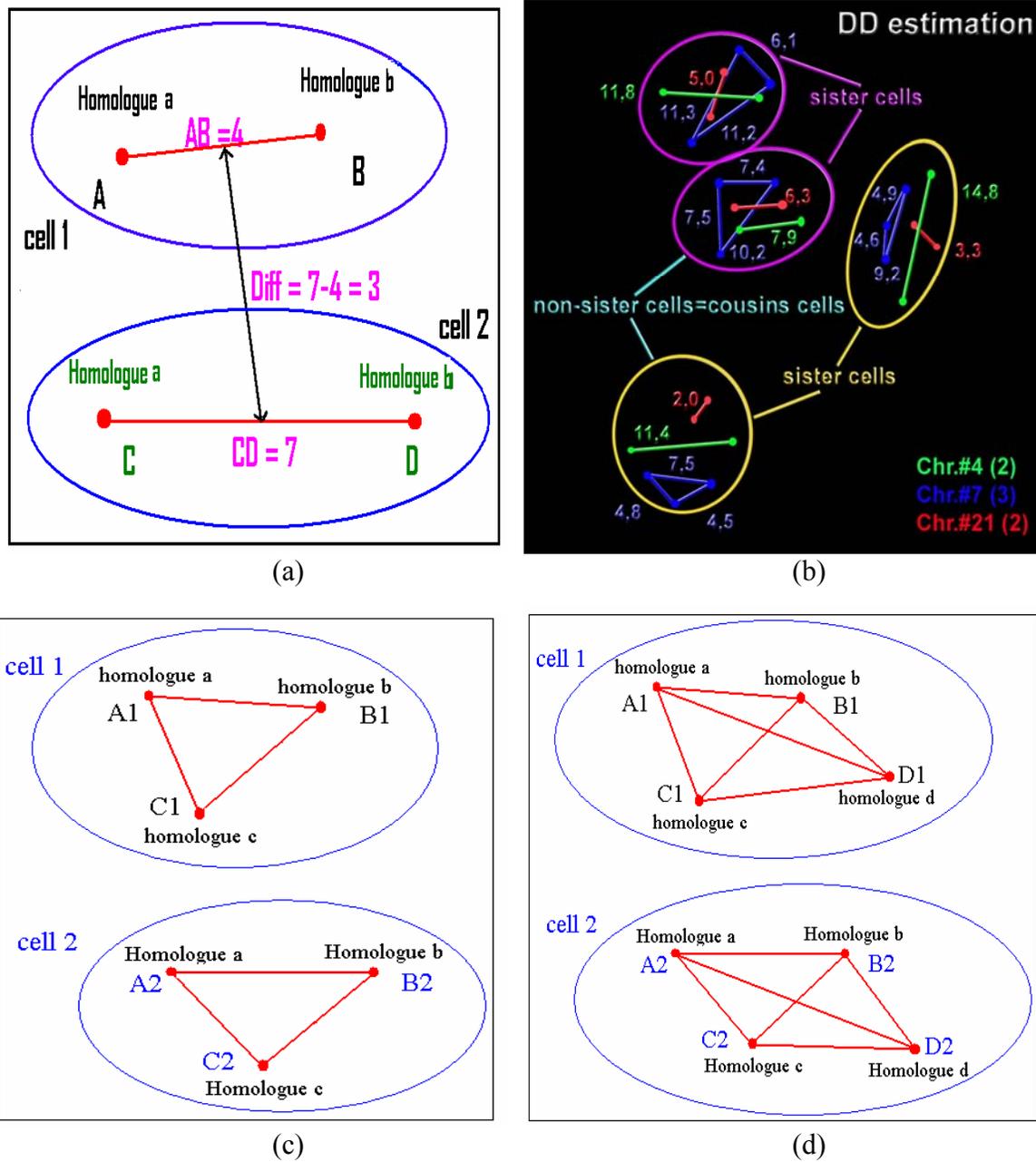


Figure 2.4 Distance Difference (DD) calculation of two homologues in one chromosome of two cell nuclei. (a) DD calculation in the case of two homologues in every chromosome. (b) DD calculation in the case of three homologues in one chromosome (chromosome 7 in blue, chromosome 4 in green and chromosome 21 in red) (Köhler et al. 2004 (B)). (c) Another example to show the DD calculation in the case of three homologues in one chromosome. (d) DD calculation in the case of four homologues in one chromosome. Note that in the data used in this thesis, no chromosome had four homologues. This case is designed for the computation in the future.

2.3.5 Angle calculation

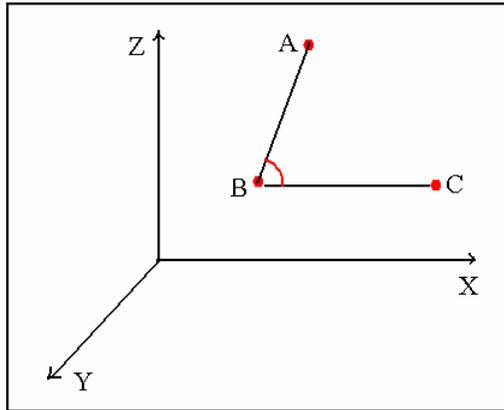


Figure 2.5 Angle calculation among point A, B and C within the Euclidean space.

In the Euclidean space, the angle θ between two vectors u and v is related to their dot product and their lengths by the formula (Harris et al. 1998):

$$\mathbf{u} \cdot \mathbf{v} = \cos(\theta) \|\mathbf{u}\| \cdot \|\mathbf{v}\|$$

In **Figure 2.5**, three red points A (a_x, a_y, a_z), B (b_x, b_y, b_z), C (c_x, c_y, c_z) are shown, which represent in our data three geometric centers of spatial entities, respectively. Vector \overrightarrow{AB} is defined as $(b_x - a_x, b_y - a_y, b_z - a_z)$, and vector \overrightarrow{BC} is defined as $(c_x - b_x, c_y - b_y, c_z - b_z)$. $|\overrightarrow{AB}|$

is the length of vector \overrightarrow{AB} . Using $\cos \angle ABC = \frac{\overrightarrow{AB} \cdot \overrightarrow{BC}}{|\overrightarrow{AB}| \cdot |\overrightarrow{BC}|}$, we can calculate the angle

between \overrightarrow{AB} and \overrightarrow{BC} . For the visualization of angle calculation see **Figure 2.6** (b). Angle calculation and 3D visualization of chromosomes in HMEC cell nuclei (Figure 2.6(a)) are shown in this figure. In Figure 2.6 (c) a detailed view of the upper right cell nucleus in (b). Figure 2.6 (d) shows another example of angle calculation for the gene *MLL* and its translocation partner *AF9*. In (d), the left image shows the 3D surface reconstructions of the gene *MLL* and its translocation partner *AF9*. The middle one shows the extracted surface points extracted from the surfaces of these two genes in the left image. The right image illustrates the angle calculation and visualization of gene *MLL* and its translocation partner *AF9*.

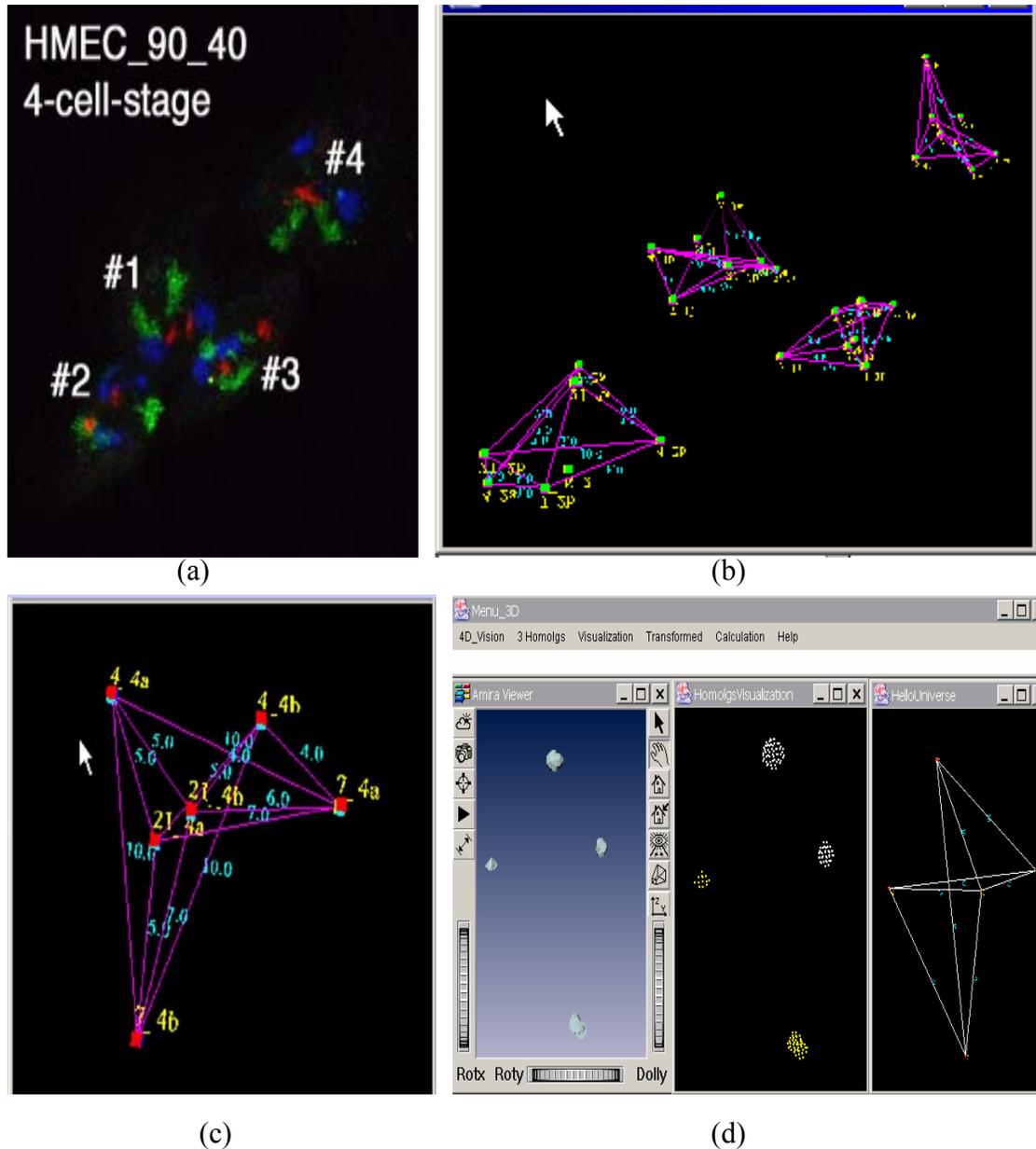


Figure 2.6 Angle calculation and related 3D visualization of chromosomes in HMEC or hematopoietic cell nucleus. (a) The project of the confocal image stack. Green: chromosome 4, blue: chromosome 7, red: chromosome 21. (image courtesy of D. Köhler). (b) Angle calculation of cell nucleus shown in (a). Green: the geometric center of chromosomes shown in (a). Pink: the distance between 2 chromosomes. (c) Enlarged visualization of the cell nucleus 4 shown in (a) and (b). Red cube: geometric center of the chromosome. (d) 3D angle calculation of gene *MLL* and its translocation partner *AF9* in hematopoietic cell nucleus. Left: surface reconstruction of these two genes. Middle:

extraction of surface points of gene *AF9* (yellow) and gene *MLL* (white). Right: distance and angle calculation of two genes.

2.4 Analysis of landmark data by the means of geometric transformations

In order to assess shape changes or the similarity of spatial arrangements, an approach is to determine the spatial transformation which maps one shape or arrangement onto the other and to investigate the properties of this transformation, in particular its “smoothness”, in the sense of producing few warping. This approach is "very much" in the spirit of D'Arcy Thomson (D'Arcy 1917) who considered the geometric transformations of one space to another (Dryden et al. 1998). If the shapes are given by two configurations X and Y (see section 2.2.3) we have to interpolate between the corresponding landmarks in order to obtain a transformation in the whole space for which smoothness can be defined. Using the thin-plate spline for interpolation ((Duchon 1976), see section 2.4.2), the smoothness of the resulting transformation can be defined by a term describing the bending energy of a thin metal plate related to this transformation. Bending energy can be defined for this transformation (Bookstein 1989). If we use the simpler rigid and affine transformations to map the two configurations onto each other, generally, corresponding landmarks can not be brought into superposition but only as close as possible together. The remaining distances can be quantified as a residual error. The latter transformations are presented in more details in the following subsections.

2.4.1 Rigid and affine Transformations

In the three dimensional space, the rigid transformation is specified by six degrees of freedom (3 rotation angles α , β , γ and 3 translations in the x, y, z direction).

$$T_{\text{rigid}}(\mathbf{x}) = \mathbf{R}\mathbf{x} + \mathbf{t}$$

Where the rotation matrix \mathbf{R} is constructed from the rotation angles α , β , γ as follows:

$$\mathbf{R} = \begin{pmatrix} \cos \beta \cos \gamma & \cos \alpha \sin \gamma + \sin \alpha \sin \beta \cos \gamma & \sin \alpha \sin \gamma - \cos \alpha \sin \beta \cos \gamma \\ -\cos \beta \sin \gamma & \cos \alpha \cos \gamma - \sin \alpha \sin \beta \sin \gamma & \sin \alpha \cos \gamma + \cos \alpha \sin \beta \sin \gamma \\ \sin \beta & -\sin \alpha \cos \beta & \cos \alpha \cos \beta \end{pmatrix}$$

The rigid transformation can accurately describe the motion of rigid structure, for instance a bone in medical imaging. But in the case of cells, this rigid transformation often results in the changes of object positions that can not simply be described using translations and rotations. The extension of this model is the affine transformation model which has twelve degrees of freedom and allows for scaling and shearing in different direction:

$$T(x, y, z) = \begin{pmatrix} a_{00} & a_{01} & a_{02} & a_{03} \\ a_{10} & a_{11} & a_{12} & a_{13} \\ a_{20} & a_{21} & a_{22} & a_{23} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

By adding additional degrees of freedom, such an affine transformation model can be extended to a non-linear transformation model and the relationship of distances can be presented in a new way.

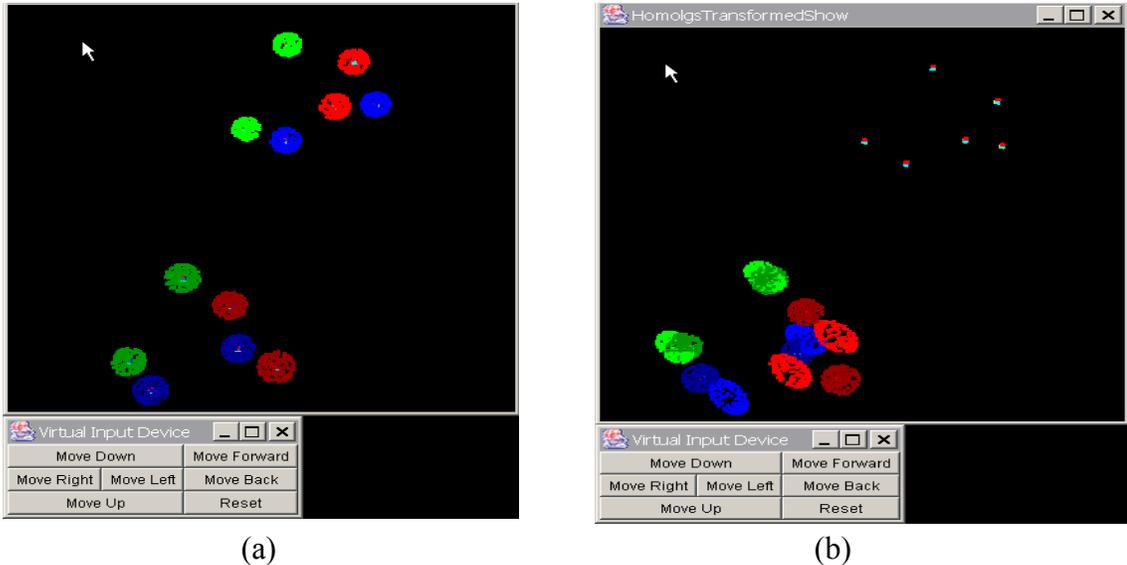


Figure 2.7 Affine transformation of HMEC cell nuclei visualized in Java 3D viewer. Green: chr. 4, blue: chr. 7, and red: chr. 21. (a) Chromosome positions in HMEC cell nucleus before affine transformation. (b) Chromosome positions in HMEC cell nucleus after affine transformation.

In **Figure 2.7** (a), every homologue is shown as a spherical point set. The homologues in light color belong to nucleus 1, and the homologues in dark color belong to nucleus 2. So

in every nucleus, there are 6 homologues. After affine transformation, as shown in (b), the positions of homologues in nucleus 1 changed and became very close to those in nucleus two. At the same time, the shape of the homologues also changed because of the affine transformation. In (b), the points at the top right show the original positions of the geometric centers of six homologues in nucleus 1. To characterize the movement of a cell or a cell nucleus the parameters of an affine transformation (mapping) have been estimated. The arrow at the top left in the figure shows that with mouse, real-time user interaction can be done in order to get a better visualization.

2.4.2 Thin-plate splines defining planar or volume transformations

Thin-plate splines (TPS) are functions that define a smooth spatial transformation mapping given corresponding points onto each other. In the original setting (Harder et al. 1972), for a given number of arbitrarily placed (pairwise different) points in the two dimensional plane, with a specified height for each point, the thin-plate spline models the surface of a thin steel plate forced to go through these points at the specified heights. Here, the spline, which is the definition of a thin-plate spline, is minimizing a quantity called integral bending norm (see section 2.4.3 for its precise definition) compared to all other functions tacking their surface to go through the same given points. This quantity corresponds approximately, under several assumptions, to the total bending energy of the steel plate. Therefore, the model is in accordance with the physical behavior of a thin steel plate which is tacked at given points and which will minimize its bending energy.

The integral bending norm $I(f(x,y))$ (or I_f) of a function f associates each point in the plane to a height value $f(x,y)$, and approximates under certain assumptions the bending energy of a thin steel plate with the surface defined by f . It is given by

$$I[f(x,y)] = \iint_{\mathbb{R}^2} f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2 dx dy.$$

For the thin-plate spline representing a two (or three) dimensional spatial transformation, the height values obtained for the transformation in x and in y (and in z in case of three dimensions) are added.

Provided n pairs $(\mathbf{p}^k, \mathbf{h}^k)$ of points $\mathbf{p}^k \in \mathbb{R}^d$ ($d = 2, 3, \dots$; $k = 1, 2, \dots, n$) and heights $\mathbf{h}^k \in$

IR, the interpolating thin-plate spline is given by:

$$f(\mathbf{x}) = a_0 + \mathbf{a}\mathbf{x} + \sum_{k=1}^n w_k U(|\mathbf{p}^k - \mathbf{x}|)$$

The scalars a_0 , w_k , and the vector \mathbf{a} are determined by the control points. U is the basic biharmonic function of the spline, $U(|x|)$ itself is proportional to $|x|^2 \log|x|^2$ in 2-D and $|x|$ in 3-D. This function is a smooth interpolation between the control points. For a comprehensive and well structured description of the algebra of TPS see Bookstein (Bookstein 1989).

The integral bending norm describes the smoothness of a transformation. As the TPS has minimal integral bending norm for a given pair (X, Y) of configuration X and Y with corresponding landmarks, we can associate this norm - and hence the smoothness - to the pair (X, Y) . The integral bending norm and its application to assess the similarity of two configurations is further discussed in section 2.4.3. Besides this property of optimal smoothness there are several benefits of TPS compared to other splines for describing spatial transformations (see also (Fieres 2001)), in particular:

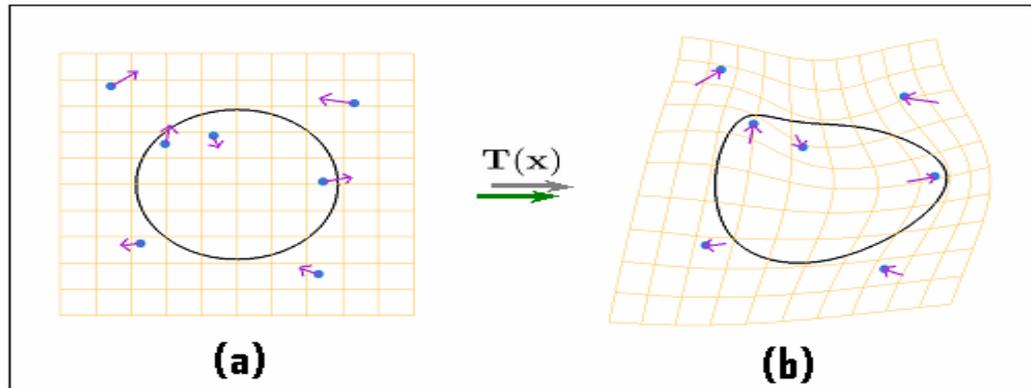


Figure 2.8 Thin-plate splines (TPS) to describe smooth deformations. The corresponding points (landmarks, blue points in (a) and (b)) are mapped onto each other as indicated by the arrows which results in the spatial transformation T shown by the yellow grid in (b).

- (1) The points in the initial configuration X (landmarks or pseudo-landmarks) are not restricted to be arranged according to a certain scheme, but are free to be set arbitrarily in space. In our case, this means that all geometric centers of chromosomes can be regarded as landmarks (Figure 2.2 (a)). One has to keep in mind, however, that the foldings of the space may appear for certain configurations and that the transformation is not invertible in this case.

(2) It is possible to model shape changes using TPS with few landmarks while obtaining a good global accuracy but preserving local features. This is useful for the automatic detection of shape correspondences with a non-rigid registration algorithm (Mattes et al. 2002). In contrast, piecewise polynomial (especially, bi- or trilinear) splines often deform also local features in order to bring two shapes with comparable remaining distance onto each other.

The corresponding points in **Figure 2.8** (landmarks, blue points in (a) and (b)) are mapped onto each other as indicated by the arrows and produce the spatial transformation T shown by the yellow grid.

2.4.3 Measuring the similarity of shapes and spatial configurations based on the integral bending norm

The value I_f should rather be interpreted as an artificial bending energy value as we are no more modeling a physical plate in the definition of the spatial transformation by means of TPS in the previous section. The integral bending norm can be calculated by a product of the form $\mathbf{w}^T K \mathbf{w}$: where K is a matrix derived from the points of the initial (landmark) configuration and \mathbf{w} is a vector formed by the weights w_k appearing in the linear combination defining the spline. The integral bending norm is dimensionless and invariant with respect to any invertible affine transformation (see section 2.4.1), especially scaling invariant. This means, for instance, a sphere of radius R has the same bending energy as a sphere of radius aR ($a > 0$). TPS have been used as morphometric tools as well as for image registration in some studies (Bookstein 1991, Mattes et al. 2002).

The bending energy overcomes the problem of cell nuclear shape and size (see section 2.3.3) and therefore it is a simplification of a quantitative evaluation suited well for comparing chromosomal arrangements and/or cell nuclear topology. As a parameter reflecting the degree of space transformation required for the best match of two chromosome territories, it is particularly well suited for the spatial arrangement evaluation of chromosomes. For the first time, BE is applied into the field of chromosomal spatial arrangement evaluation. In **Figure 2.9**, the TPS transformation from which the integral bending norm is derived is applied to the mapping of chromosome homologues (their geometric centers) to each

other.

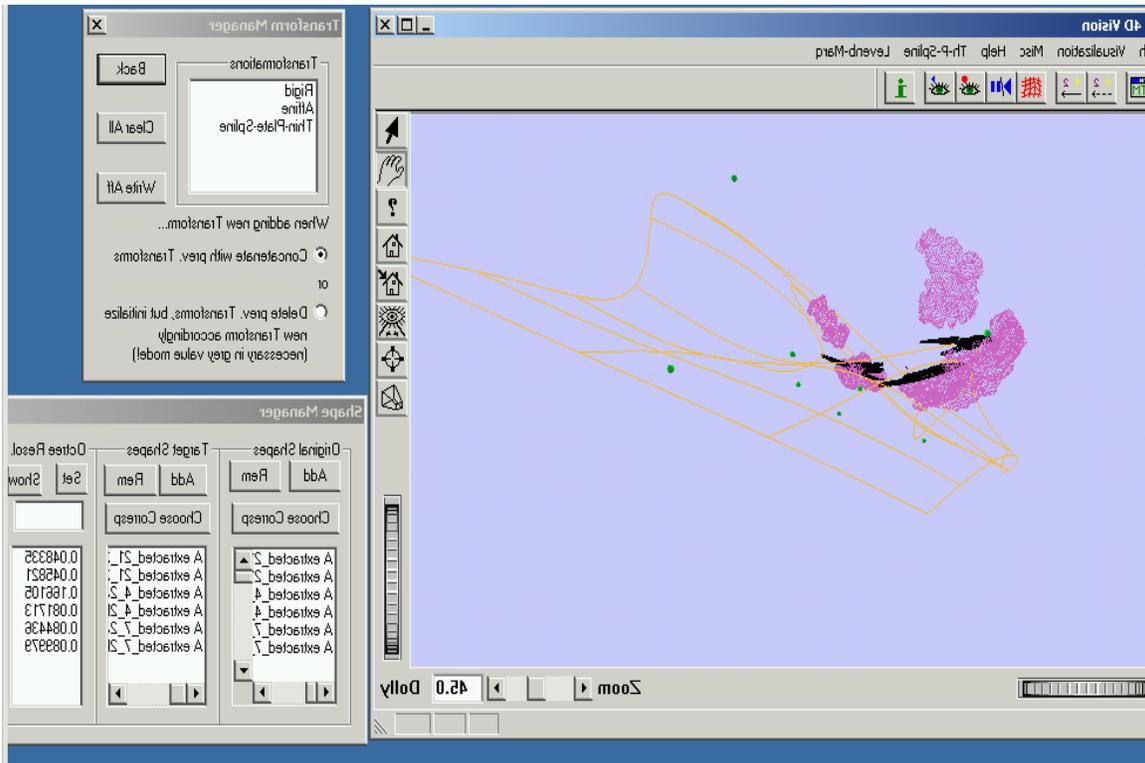


Figure 2.9 The program interface of program 4D-Vision (Mattes et al. 2002) shows the visualisation of homologues in two cell nucleus when the BE is calculated. Red: chromosome homologues in one cell nucleus. Black: The transformed chromosome homologues in the second cell nucleus. Yellow: The lattice which shows the transformed space description of the second cell nucleus.

2.5 Quantitative analysis and visualization of cell nuclear structures

2.5.1 Adaptation of the landmark based methods to the investigated data

So far, labelled landmarks have been considered, in other words we supposed that the correspondence between all landmarks of two configurations is known. However, if the two configurations X and Y represent the geometric centers of the chromosome territories in two different cells, this will not be the case, but we know the correspondence is only up to the different homologues of a chromosome. To avoid this problem, the strategy

considered in this work is to test all possible combinations of homologue correspondences and to take the minimum DD value among these combinations, i.e., the combination for which we have the best fit according to the DD value.

2.5.2 Statistical analysis

Statistical packages, such as SAS, make routine data analysis relatively easy, but they are not free software, and they make it difficult to add to the built-in capabilities of the package. In contrast, R language is a good choice to do statistical calculation for large amount of data. R language provides a good statistical computing environment which makes routine data analysis easy, and additionally, it supports convenient programming. For the R language there exist literally dozens of freely available statistical libraries of R programs ("R packages"). As well, R is particularly capable in the area of statistical graphics.

2.5.3 Visualization

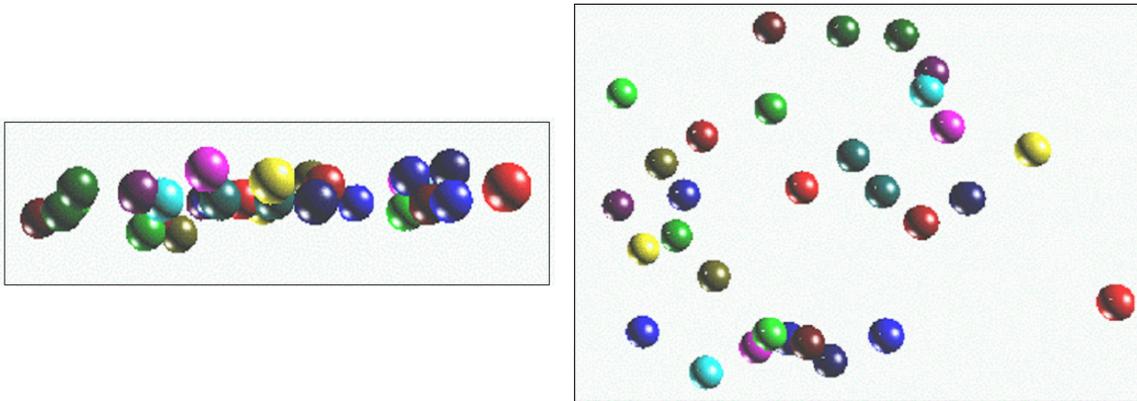


Figure 2.10 The 3D visualization of metaphase chromosome distribution using Open Inventor for Java 4.0. Every ball presents the geometric center of one homologue. The balls with the same color indicate the homologues which belong to the same chromosome. Left: the visualization through XY ortho-section. Right: the visualization through XZ ortho-section.

Commercial software like Open Inventor, and free 3D visualization tool kits like VTK (Visualization Tool Kit) or Java3D provide a detailed document (API), applications and 3D visualization. However, Java3D and its class library provide a simpler interface than most other graphics libraries, but have enough capabilities to produce 3D visualization and animation. At the same time, Java3D can incorporate objects created by 3D modeling packages like VRML models. As all computational tools presented in this thesis are written in Java 1.3/1.4, it is convenient to use Java3D and the Java3D viewer. Java3D viewer is a multidimensional virtual reality viewer that allows real-time user interaction for the visualization tasks in this thesis. As an addition to Java for displaying 3D graphics, Java 3D builds on existing technology such as DirectX and OpenGL. The Java3D package, its class library and tutorials are available free from Sun Microsystems at www.java.sun.com. For other useful web sites about Java3D see (www.java3d.org, www.j3d.org, <http://science.nasa.gov>). Another tool for 3D visualization is based on OpenInventor Java library, for an example of the 3D visualization of this library see **Figure 2.10**.

2.5.4 Aspects of implementation

In general, the interface of the program that we developed is shown in **Figure 2.11**. In order to design model-based biological experiments or experiment-based model, a new method is presented here:

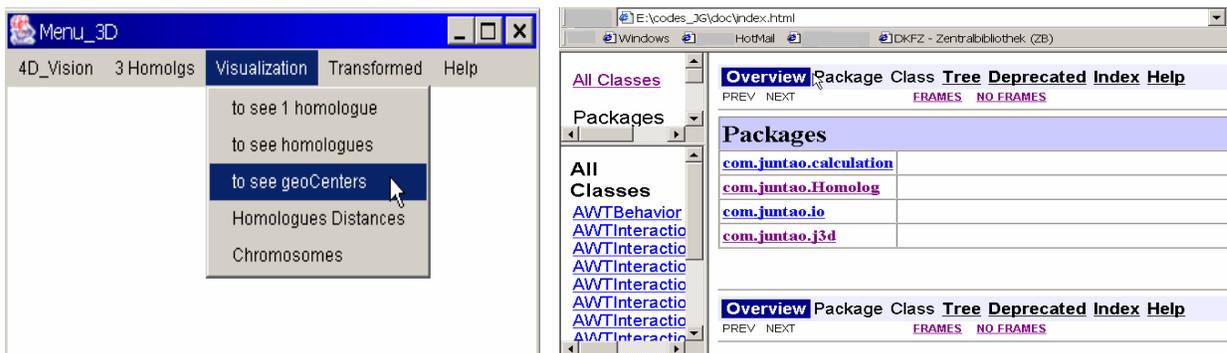


Figure 2.11 The program KanKan for quantitative evaluation used in this thesis. (a) One of the windows of the KanKan program. (b) Documentation of Java class library used in KanKan program.

- (1) The prerequisite to use this new method is that the 3D surface reconstruction of data sets based on confocal image stacks has been obtained at first.
- (2) Based on the 3D surface reconstruction, the geometric centers are calculated. This can be done automatically using the computational tools presented in this thesis.
- (3) Based on the geometric centers, the distances, distance difference, angles and Hausdorff distances can be calculated. How to realize the automation of these calculations depends on the structure of data (chromosome, replication loci, gene, or other nucleus entity).
- (4) The quantitative methods from statistics (for example, Wilcoxon Signed Rank test, Kolmogorov-Smirnov test, etc.), image processing and computer vision (for example, thin-plate splines transformation, bending energy, etc.) can be applied thereafter.

2.5.5 Principles of Object-Oriented Programming (OOP)

OOP is a software design method that models the characteristics of abstract or real objects using classes and objects (Kay 1993). A class is a prototype to define the variables and the methods common to all objects of a certain kind. Software objects interact and communicate with each other by sending messages to each other. Like other OOP languages, Java is based on the five principles of OOP (Kay 1993, Budd 2001). The data in this thesis is organized according to these OOP principles (Eckel 2002), as shown in **Figure 2.12**. In this figure, there are several classes which are Clone, Cell and Homologue, respectively. In every class, there are characteristics (e.g. Name) and behaviors (also called methods, e.g. getName()). The structure of Homologue, Cell and then Clone is an analogy to the objects in the real biological world.

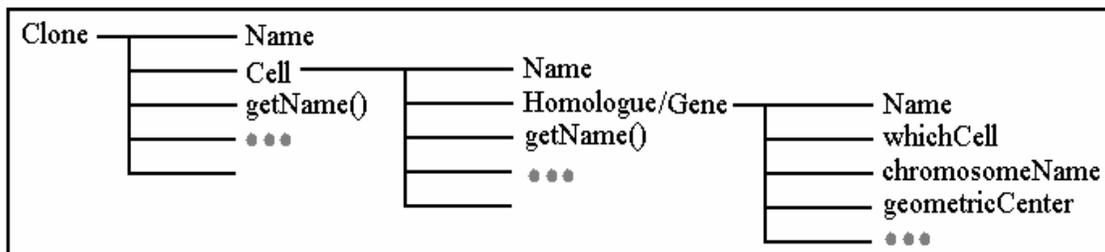


Figure 2.12 Data organization under OOP principles.

Chapter 3 Quantitative analysis of the inheritance and changes of chromosome arrangement in proliferating human cells

3.1 Introduction

3.1.1 The chromosome position similarity in daughter nuclei

Theodor Boveri suggested in 1909 that, in the nematode *Parascaris Equorum*, chromosomal neighborhoods established in the metaphase plate of blastomere nuclei are maintained during anaphase and telophase in the two separated sets of chromatids, which results in a rather symmetrical Chromosome Territory (CT) arrangement in the two daughter nuclei (Boveri 1909, Walter et al. 2003). Another study in 1932 indicated rather symmetrical locations of nucleoli and chromocenters in daughter nuclei from a variety of plant species (Heitz 1932). Recently, FISH experiments suggested a considerable degree of symmetry in the arrangement of whole CTs and centromeres in daughter nuclei (Sun et al. 1999, Habermann et al. 2001). More recent study of normal rat kidney (NRK) cells during mitosis also demonstrated the spatial symmetry of chromosome positions in daughter cell nuclei (Gerlich et al. 2003 (A)). 3D FISH experiments with chromosome paint probes in HeLa cell clones at two- and four-cell stage suggested that there is a potential symmetry in daughter cell nuclei, but major difference of chromosome arrangement can be already recognizable at the four-cell stage (Walter et al. 2003).

Although these studies show the evidence that chromosome positions in daughter cell nuclei have similarity, it is not clear how similar chromosome positions in daughter cell nuclei are and whether chromosome positions still have similarity after several cell generations. At the same time, these studies arise question: Is there any possibility to quantify the potential symmetry of chromosome positions in daughter cell nuclei? If yes, how?

3.1.2 Recent studies of chromosome arrangement inheritance

There is controversy about the extent to which chromatin organization is inherited from mother to daughter nucleus. Three studies (Gerlich et al. 2003 (A), Walter et al. 2003, Thomson et al. 2004) and very shortly thereafter five reviews (Parada et al. 2003, Bickmroe et al. 2003, Williams et al. 2003, Gerlich et al. 2003 (C), Pederson 2004) have addressed how chromosome arrangement unfolds into the subsequent interphase.

In one study (Gerlich et al. 2003 (A)), authors employed non-invasive labeling of chromosome subsets in rat kidney cells, in vivo tracking by 4D imaging, together with computer simulation and prediction. They have concluded that the interphase global arrangement of chromosomes remains relatively constant and is inheritable through the whole cell cycle in mammalian cells. In another study (Walter et al. 2003), authors used long-term live-cell studies of HeLa cells with GFP-tagged chromatin and showed that CT arrangement was stably maintained from mid G1 to late G2/early prophase, whereas major changes of CT neighborhoods occurred from one cell cycle to the next one. In the third study (Thomson et al. 2004), time-lapse microscopy was employed to show that the association of loci with nuclear compartments displays significant asymmetry between daughter nuclei and therefore cannot be inherited from the mother nucleus.

These studies differ with respect to the time window examined: one extended their investigation until 4-cell stage (Walter et al. 2003), other examined only single cell cycle (Gerlich et al. 2003 (A), Thomson et al. 2004). Importantly, in all three mentioned studies different cell lines were used --- HeLa cell line (Walter et al. 2003), rat embryonic kidney fibroblast-like cell line (Gerlich et al. 2003 (A)) or human HT-1080 cell line (Thomson et al. 2004) --- which should be taken in mind when different results of the studies are discussed.

Despite the advantages of live cell imaging and analysis, several limitations of the methods employed by the three studies should be considered:

- (1) Only spatial arrangement of chromosomes between one generation and the next one was studied;
- (2) In every of all mentioned studies, chromosome arrangement was examined in only one cell line;

(3) Quantitative analysis of chromosome arrangement has not yet been provided.

3.1.3 New strategy to study changes in chromosome arrangement between different cell generations

The earlier studies into inheritance of chromosome arrangement were based on *in vivo* observations of chromatin labeled with fluorescing proteins after bleaching before mitosis in a certain portion of the nucleus (see above). The advantages of this approach are absence of fixation artifacts and direct observation of the changes that take place. However, this approach also has disadvantages: (1) Fast recovery of fluorescence prevents observations for more than one cell division; (2) Chromosome origin of the observed chromatin regions cannot be determined; (3) Changes may be traced only as changes of the border between labelled and not labelled chromatin fractions - what happens within labelled and not labelled areas remains hidden and (4) comparison to not-related cells is not possible.

An alternative approach that has not been used so far is based on visualization of chromatin using FISH. Though FISH causes certain damage of nuclear architecture due to fixation and denaturation of DNA, the observed chromatin shift does not affect chromosome arrangement (Solovei et al. 2002). The main disadvantage is that direct continuous observations are impossible. One has to estimate the changes in chromosome arrangement indirectly, based on the diversity of chromosome arrangements accumulated within clones of different age. This drawback is compensated by a number of important advantages: (1) Changes in chromosome arrangement may be studied for several successive cell divisions; (2) One can exactly trace changes in the positions of specified chromosomes, (3) Knowing the positions of identified chromosomes allows us to use methods that are specifically sensitive to the shift of chromosome in relation to one another and are therefore especially promising for studies of chromosome arrangement, and (4) the differences observed within clones may be compared with the golden standard given by not-related cells.

3.2 Data description

In order to quantify the inherity of chromosome spatial distribution, it is necessary to select the suitable cell lines which can be cultured and fixed easily after several cell divisions. Seven human cell lines were tested but finally only four cell lines have been chosen because it was easy to keep track of their clonal growth after cell divisions (Köhler 2002). These four cell lines are: Two normal human cell lines HMEC (human mammary epithelial cell) and HFb (Human Fibroblast), two cancer cell lines HeLa and DLD. As the first permanent cell line, HeLa can be cultured and fixed easily. In HFb cell line, there are two kinds of cells (see Table 3.1). One is called HFb_CB, which stands for Cytochalasin B, a reagent to let nuclei divide but prevent cells dividing as HFb cells move a lot. The other kind of cells is called HFb Ki67 (Ki-67 is a nuclear protein to detect the cell cycle stage of the cells).

3.2.1 Cell culture and the confocal images

Cells were grown on “gridded” coverslips to keep track of cells during clonal cell growth (**Figure 3.1**). Cells from four different cell lines were monitored every day and fixed at different stages of clones growth: 2-cell, 4-cell, 8-cell, 16-cell, 32-cell stages. Fixation with 4% formaldehyde and necessary pretreatments for 3D-FISH was performed according to a standard protocol (Solovei et al. 2001). Only clones which could be traced back to a single cell were used for further analysis (Köhler et al. 2004 (A)).

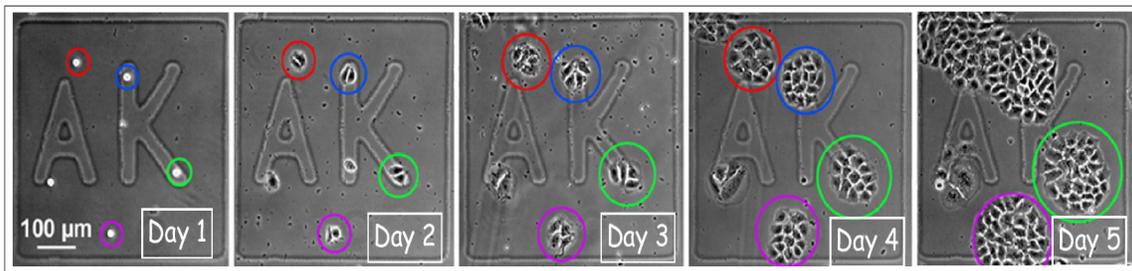


Figure 3.1 The clone growth at difference stages. Day 1: 1-cell stage, day 2: 2-cell stage, day 3: 4-cell stage, day 4: 8-cell stage, day 5: 16-cell stage (Köhler et al. 2004 (B)). The letter "AK" is to mark the different areas on one coverslip, it is not related to the clone growth directly.

Only three chromosomes have been stained according to the three different color channels of confocal microscope. As only fixed cells could be stained, in each experiment the last generation of cells has been imaged.

After the FISH procedure, clones were identified on a Leica TCS SP1 confocal microscope using a Plan Apo 25 x /1.0 NA oil immersion objective. Then, the whole nuclei series of light optical sections were collected using Plan Apo 100 x /1.4 NA oil immersion objective. Small clones with 2 - 4 cells were collected as one image stack; for larger clones a few image stacks were collected. For each optical section, images were collected sequentially for all three fluorochromes mentioned above. Fluorochromes were visualized using an argon laser with the excitation wavelengths of 488 nm (for Alexa 488) and 514 nm (for Cy3), and a helium-neon laser with the excitation wavelength of 633 nm (for Cy5). Stacks of 8-bit gray-scale images were obtained with axial distances of 400 nm between optical sections and pixel sizes ranging from 50 to 500nm depending on selected zoom factor. Galleries of confocal RGB images were assembled using ImageJ and Adobe Photoshop programs. Three-dimensional reconstructions of chromosome territories were performed by surface segmentation of image stacks using Amira 2.3 software (<http://www.amiravis.com>). For the confocal images and their 3D surface reconstruction see **Figure 3.2** and **Figure 3.3**.

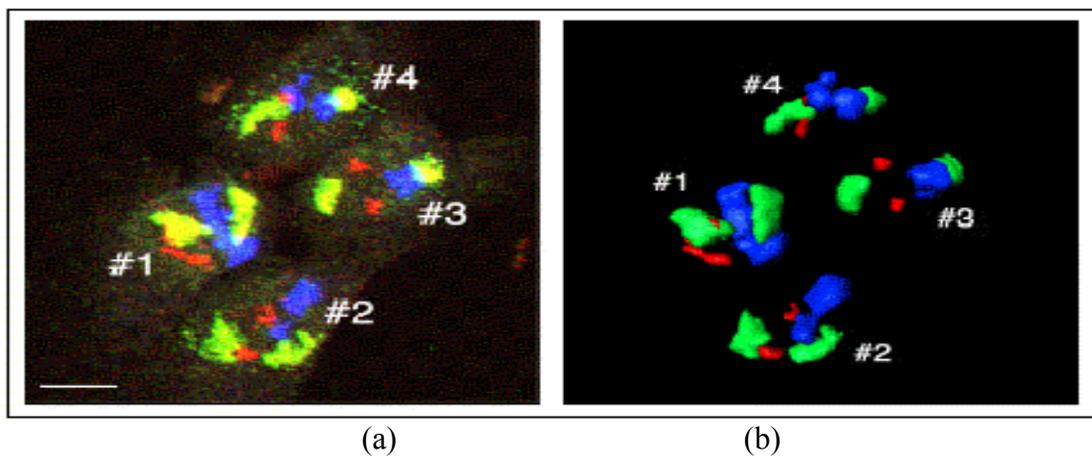


Figure 3.2 Chromosome 4, 7 and 21 arrangements in HMEC cell #1, #2, #3 and #4 of one 4-cell-stage clone. (a) Projections of confocal image stack obtained after painting of chromosome 4 (visualized in green), 7 (visualized in blue) and 21 (visualized in red). (b) 3D surface reconstruction of chromosomes (image courtesy of D. Köhler).

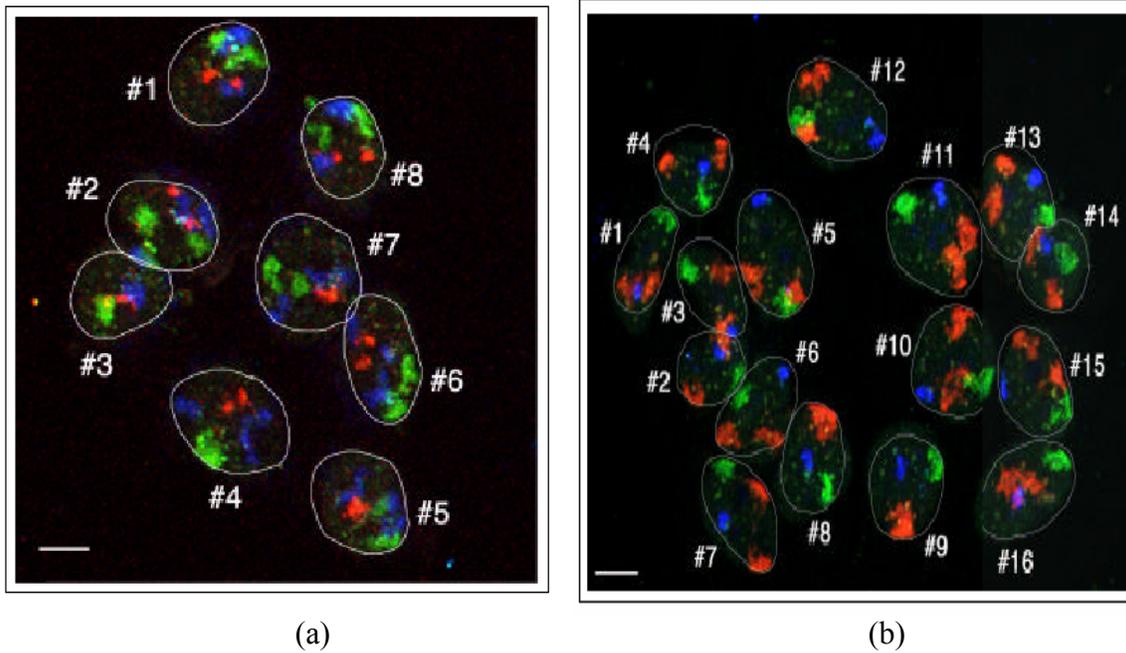


Figure 3.3 Chromosome arrangement in 8-cell-stage and 16-cell-stage clones of different cell lines. (a) Chromosome 4 (green), 7 (blue) and 21 (red) arrangements in one HeLa 8-cell-stage clone. (b) Chromosome 7 (red), X (green) and Y (blue) arrangements in one DLD 16-cell-stage clone (image courtesy of D. Köhler).

3.2.2 Different cell generations

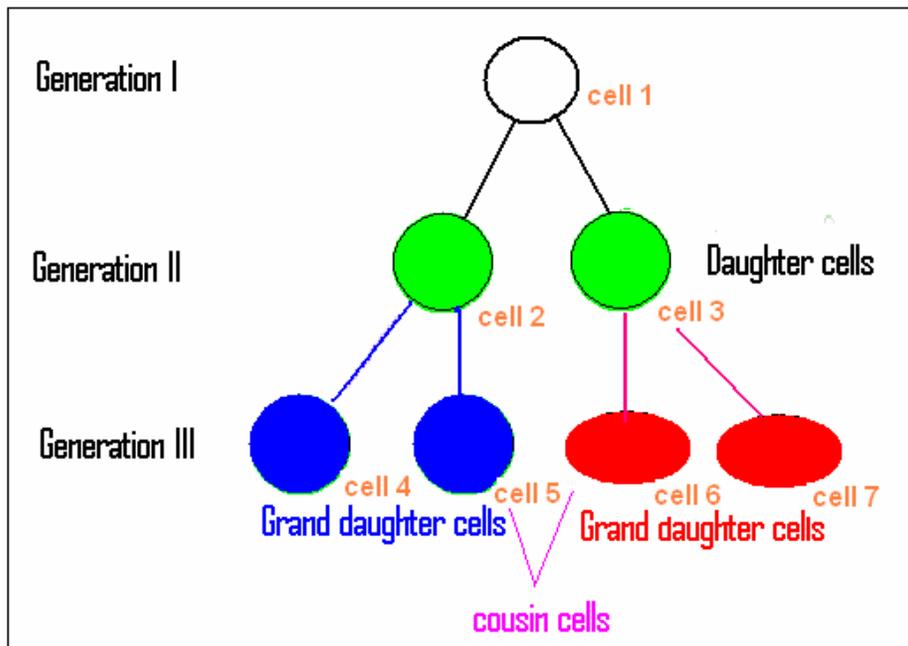


Figure 3.4 The illustration of sister cells, cousin cells in three cell generations.

Figure 3.4 is a diagram showing the three cell generations in one cell line. In generation I, there is one mother cell. Generation II, which is also called 2-cell stage, consists of two daughter cells obtained from the mother cell. Generation III, called 4-cell stage, are the four cells obtained from the two daughter cells. Similarly, generation IV (8-cellstage), V (16-cellstage), and VI (32-cellstage) were obtained (not shown here). Sister, cousin, and non-related cells were defined as following: (1) Sister cells: the cells generated from the same mother cell. In Figure 3.4, sister cells are the cells that have the same color. For example, cell 2 and 3, cell 4 and 5, cell 6 and 7 are sister cells, respectively. (2) Cousin cells: cells who's mothers are sisters. For example, cell 5 and 6, cell 5 and 7, cell 4 and 7 are cousin cells. (3) Non-related cells: cells without common ancestors. The cells in one RCC (Randomly Chosen Cells from growing population) clone can be regarded as non-related cells, see section 3.2.4 for the details.

3.2.3 Data description of the four cell lines

Table 3.1 The data description of four cell lines. Note: cst means cell stage.

Cell line		The stained chromosome	Number of cell generations	Number of cell clones	
HeLa		# 4 (green) # 7 (blue) # 21 (red)	5	2cst	22
				4cst	18
				8cst	5
				16cst	3
				32cst	2
HMEC		# 4 (green) # 7 (blue) # 21 (red)	2	2cst	16
				4cst	24
HFb	CB	# 4 (green) # 7 (blue) # 21 (red)	1	2cst	22
	Ki67	# 7 (red) # X (green) # Y (blue)	1	2cst	10
DLD		# 7 (red) # X (green) # Y (blue)	5	2cst	13*
				4cst	14
				8cst	5
				16cst	3
				32cst	2

* clone DLD_20 is excluded because the 2 homologues of chromosome 7 are not divided.

Totally four cell lines were chosen for the quantitative evaluation. For every cell line, there are several cell generations. In every cell generation, there are several cell clones. In every cell clone, there are different number of cells, which are sister cells, cousin cells or non-sister cells, respectively. The cell number changed from 2 to 38. In order to understand the whole data in a better way, it is necessary to get one table to describe all data produced by confocal microscope. The cells in these four cell lines are presented in Table 3.1. In this table, HeLa and DLD cell lines can be monitored easily, so they have five generations, respectively. But for HMEC and HFb cell line, it is hard to keep track of cells during their growth. In this case the monitoring was stopped earlier (Köhler 2002).

3.2.4 RCC clone definition and its realization

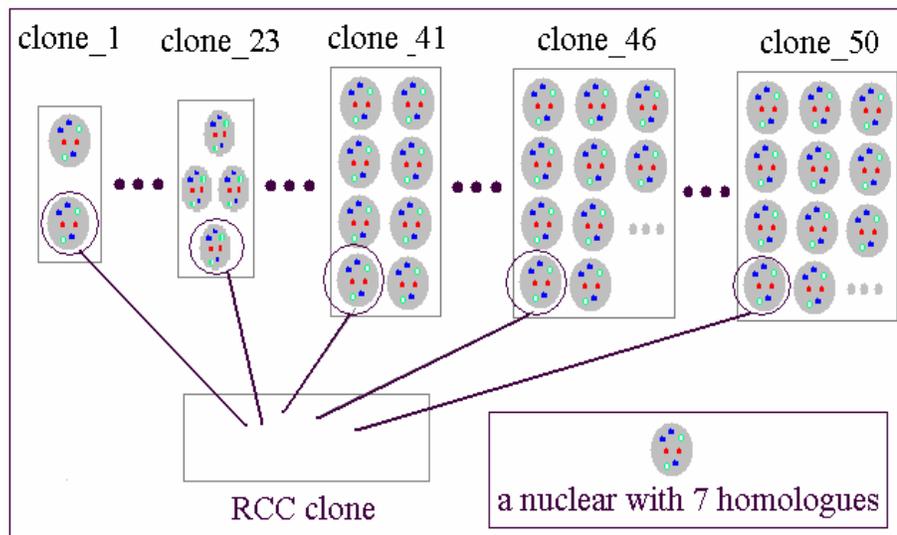


Figure 3.5 The illustration of RCC clone in HeLa cell line.

Based on the well-monitored four cell lines, we are able to compare the spatial similarity of cell nuclei through every cell generation. However, the cells in every generation are related to each other and have some relationships: they are sisters or cousins. We still need some clones in which the cells are basically unrelated and separated by several cell divisions. To collect such data straightforward would mean a lot of scanning which can not be done technically at the moment. However, with the help of computer, we can choose one cell from every clone randomly, and put these cells into a new "artificial

clone", which is called RCC clone (Randomly Chosen Cells from growing population, **Figure 3.5**).

How to get one RCC clone from all HeLa cell clones? In the data of HeLa cell line, from 2cst to 32cst, altogether there are 50 clones (see Table 3.1). In every clone, one cell is chosen randomly using Java program. Finally one HeLa RCC clone is obtained which has 50 cells chosen from 50 clones respectively. As just one cell is chosen randomly from one clone, the 50 cells are unrelated to each other. The Bending Energy (BE) and Distance Difference (DD) of RCC clones were calculated and compared with the values of the original clones in every cell generation, for the result of these RCC clones see section 3.4.

3.3 Quantification of the similarity of CT spatial arrangement

3.3.1 BE calculation depends on the setting of landmarks

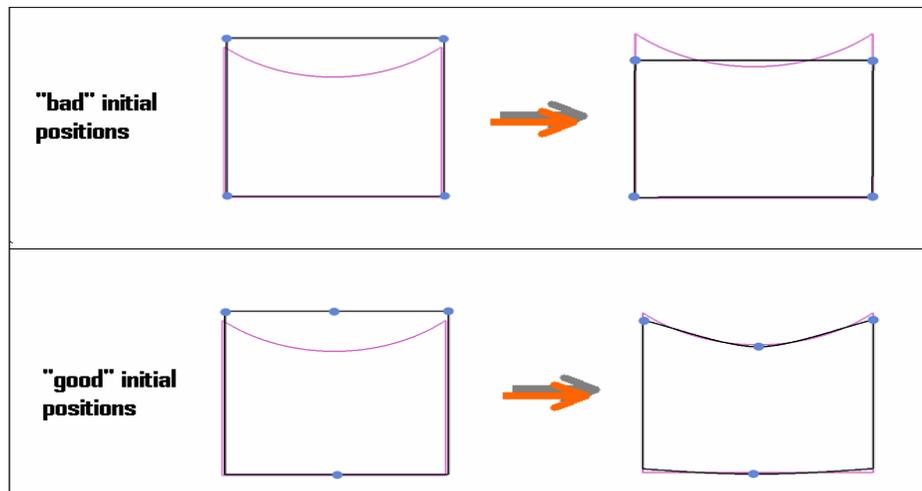


Figure 3.6 Registration of contours in two different ways to place the landmarks (blue points). Left column: Sensed (black) and reference (purple) shape with two different landmarks initializations. Right column: matching result. In the upper picture the necessary deformation can not be obtained, whereas a fair match is possible in the lower example. Note that the number of control points in both examples is equal.

As a technique for the evaluation of spatial symmetry, the calculation of Bending Energy depends on the setting of landmarks (**Figure 3.6**). The positions of the landmarks have a strong impact on the motion model and hence on the match and the values of BE (Mattes et al. 2002). For the data in this thesis, the geometric center of one homologue is regarded as one landmark.

3.3.2 The landmark setting in four cell lines

The landmark setting is the key step for the BE calculation in HeLa, HMEC and HFb_CB cell line. As shown in Table 3.2, in every HeLa cell, there are 7 homologues, so there are 7 landmarks; in every HMEC cell or HFb_CB cell, there are 6 homologues, so there are 6 landmarks.

In HFb Ki67 and DLD cell line, there are only 4 homologues (or landmarks) which were not enough to calculate the BE using thin-plate spline interpolation, so PCA-normalized spheres (or random spheres) were used instead of landmarks.

Table 3.2 The landmark number in the cell nucleus of different cell lines.

Cell line		Stained Chromosome	Homologue number in every chromosome	The number of landmarks
HeLa		## 4	2	7
		## 21	2	
		## 7	3	
HMEC		## 4	2	6
		## 21	2	
		## 7	2	
HFb	CB	## 4	2	6
		## 21	2	
		## 7	2	
	Ki_67	## X	1	4
		## Y	1	
		## 7	2	
DLD		## X	1	4
		## Y	1	
		## 7	2	

To calculate the BE, two point sets are necessary: one is reference point set which stays constant while the sensed point set is transformed, the other is sensed point set which is

transformed to resemble the reference point set. This means, for every BE calculation, two cells should be chosen. In order to set the landmarks and to calculate the BE in a correct way, two kinds of combinations in every clone have been considered: cell combinations and homologue combinations.

3.3.2.1 Cell combinations in every cell clone

To do the matching, every time two cells have to be chosen from one clone of one cell line arbitrarily, as the cells are named randomly in every clone. So it is necessary to get all combinations of all cells in every clone. Given one clone which has n cells, there will be $C_n^2 = n \times (n-1) / 2$ matchings for the cell combinations in this clone. For example, in one 8-cell clone, there will be $C_8^2 = (8 \times 7) / 2 = 28$ matchings.

3.3.2.2 Homologue combinations in every two cells

In every cell nucleus, the homologues of every chromosome are named randomly. As shown in **Figure 3.7**, homologue a, b or c in every chromosome is named randomly, and we do not know exactly which homologue is a, and which homologue is b. This makes it more difficult to do the BE calculation, as we do not know that the homologue a of chromosome 4 in nucleus 1 is corresponding to the homologue 4a or 4b in nucleus 2. For the chromosome 7 or chromosome 21, we do not know neither. This rises the question: for every landmark in nucleus 1, how to choose the correspondent landmark in nucleus 2.

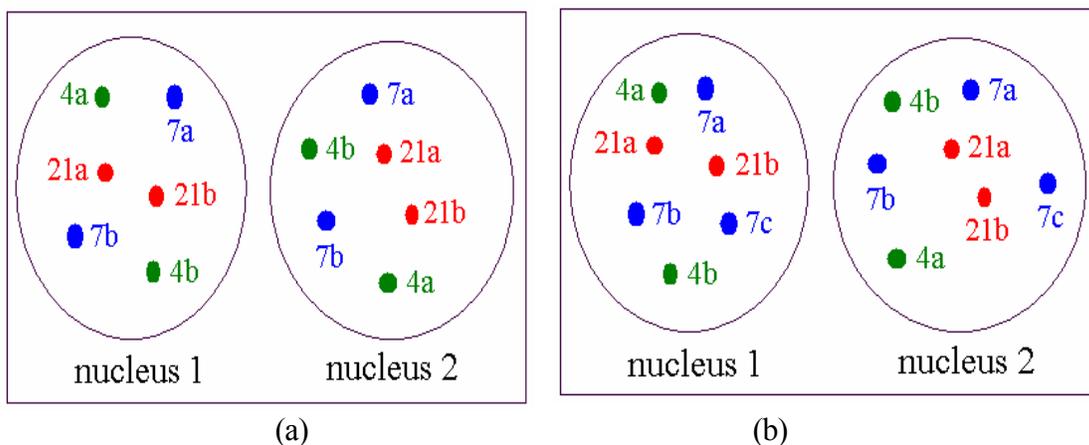


Figure 3.7 The spatial distribution of chromosome homologues. Every point, as the geometric center of one homologue, is considered as one landmark when Bending Energy is calculated.

Red: chromosome 21, Green: chromosome 4, Blue: chromosome 7. "4a" stands for the homologue a in chromosome 4. (a) There are altogether 6 homologues in HMEC cell (2 homologues in chromosome 7). (b) There are altogether 7 homologues in HeLa cell (3 homologues in chromosome 7).

(1) HMEC cell line

In the cell nucleus 1 of Figure 3.7 (a), the homologue **a** in chromosome 4, as shown as 4a in the Figure 3.7 (a), has two possibilities to find the corresponding homologue (or landmark) in cell nucleus 2: either homologue **a** of chromosome 4 in nucleus 2 (This matching can be written as 1_4a:2_4a, in which "1_" stands for cell 1, "4a" means the homologue a in chromosome 4), or the homologue **b** of chromosome 4 in nucleus 2 (this matching will be 1_4a:2_4b).

The matching in chromosome 21 which has also 2 homologues is similar as chromosome 4: it has also two possibilities to get the corresponding landmark in cell nucleus 2 (1_21a:2_21a or 1_21a:2_21b). Similarly, chromosome 7 has two possibilities to find the corresponding landmark in cell nucleus 2 (1_7a:2_7a, 1_7a:2_7b) as well. So for the three chromosomes in HMEC cell nucleus 1, there are $2 \times 2 \times 2 = 8$ possibilities to get the corresponding landmarks in cell nucleus 2.

(2) HeLa cell line

But in every HeLa cell, there are three homologues in chromosome 7. As shown in Figure 3.7 (b), chromosome 4 and 21 have two possibilities to get the corresponding landmarks, respectively. But chromosome 7 has $3 \times 2 \times 1 = 6$ possibilities to find the corresponding landmarks in nucleus 2. So for the three chromosomes in HeLa cell nucleus 1, there are $2 \times 2 \times 6 = 24$ possibilities to get the corresponding landmarks in cell nucleus 2.

For the homologue combinations between every two cells, there will be 8 (in the case of HMEC cells) or 24 (in the case of HeLa cells) possibilities to get the corresponding landmarks. This means that 8 or 24 BE values should be obtained for the matching of every two cells. The minimum of these 8 or 24 values is chosen as the characteristic BE for the matching of these two cells.

3.3.2.3 The calculation of BE

Once the characteristic BE for the matching of every two cells is obtained, all cell combinations can be taken into account. For example, in one 8-cell HeLa clone, there are 28 cell combinations; in every cell combination, there are 24 homologue combinations. The minimal BE is chosen from these 24 homologue combinations as the characteristic BE in every cell combination. For 28 cell combinations, 28 characteristic BEs are obtained, and the arithmetical mean of these 28 Bes is regarded as the BE of this clone. Using cell combinations and homologue combinations, BE can be calculated for every clone. The BE calculation by this way corresponds well to the intuitive visual inspection of the biologists, see section 3.4.2 for the details of the identification of sister and cousin cell pairs in four-cell clone.

3.3.3 Codes improvement to calculate the BE

The methods introduced in 3.3.2 to set the landmarks and to calculate the BEs were applied successfully and the results are shown in section 3.4.1. These results show that BE method is sensitive enough to quantify the large amount of data in order to describe the CT distributions during different cell generations. However, the amount of the files produced to set the landmarks is huge. For example, in the case of one HeLa 32-cell clone, there will be $C_{32}^2 = (32 \times 31) / 2 = 496$ matchings, and in every matching, there are 24 files to set the landmarks. Altogether there will be more than 10,000 files produced to calculate the BE in just one cell clone. The production of this amount of files is time-consuming and new codes are expected to be written.

Based on the calculation mentioned above, new codes written in Tcl/Tk by Dr. J. Mattes were provided which can produce the landmark files and calculate the BE in a more efficient way. Comparing to the result in **Figure 3.8** obtained using java codes, the new result obtained using Tcl/Tk codes keeps the same values and the same conclusion, that is, the spatial arrangement of CTs in sister nuclei was more similar compared to (4cst, 8cst or) cousin nuclei, and the dissimilarity of the CT arrangements during different cell generations increased monotonously.

3.3.4 PCA-normalized spheres of DLD cell line

There are only four homologues in every DLD cell or HFb_Ki67 cell, which means that there are only 4 landmarks which are not enough for the BE calculation. Based on the geometric center of every homologue, the randomly produced sphere or PCA-normalized (Principle Component Analysis) sphere around this geometric center was produced. After the transformation, the shortest distance between every point of the homologue sphere in one cell and the whole sphere point set of another cell was calculated to evaluate the similarity of CT arrangement in every two cells.

3.4 BE and DD results

3.4.1 BE result in several cell lines

3.4.1.1 HeLa cell line

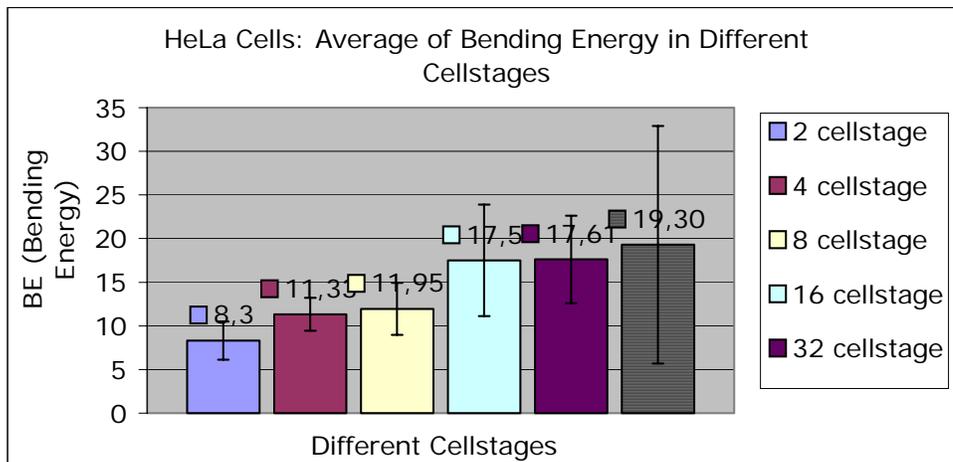


Figure 3.8 Bending Energy calculated for the different cell stages of HeLa cell line.

In Figure 3.8, BE value of HeLa 2-cell stage (8.3) was smaller significantly than the BE of HeLa 4-cell stage (11.33). This means that the spatial arrangement of CTs in 2-cell-stage cell nuclei was more similar compared to 4-cell-stage. It can be concluded that CT spatial arrangement in sister nuclei was more similar than cousin nuclei. The dissimilarity of the chromosome spatial arrangements increased monotonously. After 6 generations (32-cell stage) a value close to RCC cells was reached. RCC cells are coming from different clones

and thus reflecting the dissimilarity of all clones in the whole cell line (see the bar with the value 19.30 in Figure 3.8).

3.4.1.2 HMEC cell line

The BE result of HMEC cell line is shown in **Figure 3.9**. The BE of four-cell stage (8.68) is larger significantly than the BE of two-cell stage (5.04). This also suggests that the spatial arrangement of CTs in sister nuclei was more similar than cousin nuclei.

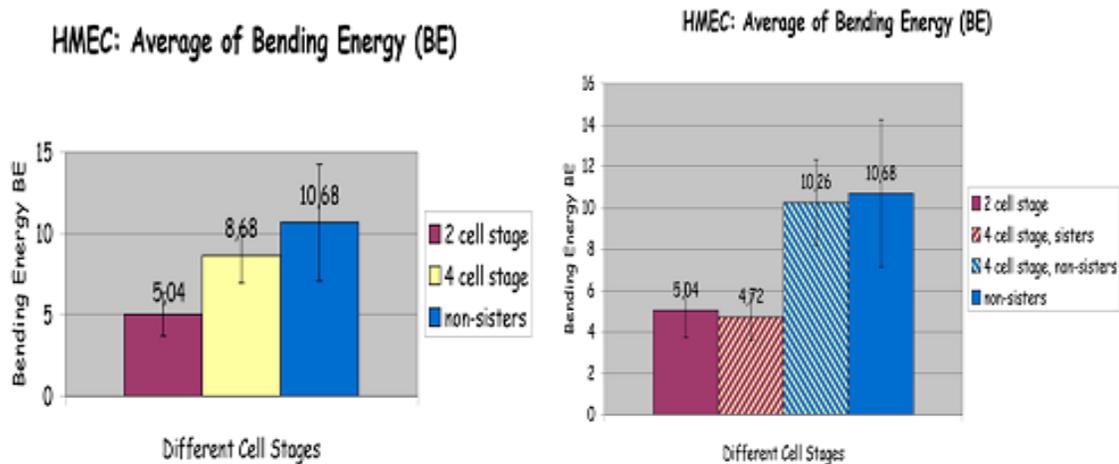


Figure 3.9 The BE result of different cell stages in HMEC cell line.

3.4.1.3 HFb_CB cell line

For HFb_CB cell line, there are only data for 2 cell stage (22 cell clones), and the BE of 2-cell stage of HFb_CB cell line calculated using the method described in section 3.3 is 3.42 ± 1.94 . This low BE value suggests that the spatial configurations of CTs in every two daughter cell nuclei are pretty similar in human fibroblast cells.

3.4.2 The identification of sister and cousin cell pairs in four-cell clones

Using the cell combinations and homologue combinations described in section 3.3.2, BEs were calculated in 4-cell clones of HMEC cell line. In Table 3.3 there are 6 BE values between every two cells of HMEC cell clone No. 25. For HMEC cell clone No. 25, one can see that cell 2 and cell 3, cell 1 and cell 4 have the smallest Bending Energy value,

respectively. As the BE for two sister nuclei is always smaller than for two non-sister nuclei, these two values predict that cell 2, 3 are sister cells, and cell 1, 4 are sister cells respectively.

Table 3.3 The bending energies of one four-cell clone in HMEC cell line.

Combinations of every two cells in HMEC clone No. 25	Bending Energy
Cell 1 and cell 2	9.653344
Cell 1 and cell 3	8.868111
Cell 1 and cell 4	4.991826
Cell 2 and cell 3	1.012165
Cell 2 and cell 4	27.542143
Cell 3 and cell 4	34.775501

In **Figure 3.10**, from the projection (a) and 3D reconstruction (b), it can be judged virtually and easily, that cell 2,3 are sister cells, and cell 1, 4 are sister cells as well. And this visualization corresponds to the calculation of Bending Energy in Table 3.3 directly.

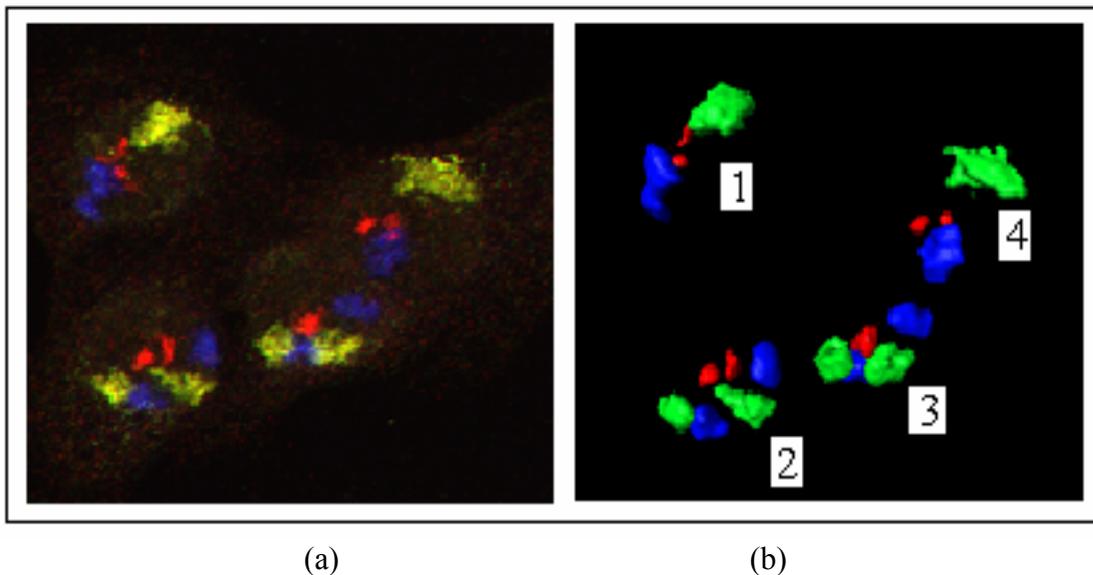


Figure 3.10 HMEC CT #4 (green), #7 (blue) and #21 (red) arrangements in nuclei of one 4-cell clone (HMEC clone No. 25) (image courtesy of D. Köhler). (a) Projections of confocal image stacks obtained after painting of chromosome #4, #7 and #21. DNA counterstain is not shown here. (b) The 3D reconstruction of CT #4, #7 and #21 arrangements in nuclei of one HMEC 4-cell clone.

Another nice example to show that BE can identify the sister and cousin cell pairs is from HeLa cell clone (Köhler et al. 2004 (B)). In the four cells of **Figure 3.11(a)**, it is difficult to judge visually if cell 1, 2 are sister cells, or cell 1, 3 or cell 1, 4 are sister cells. With 3D FISH technology in Figure 3.11 (b) and BE quantification in Figure 3.11 (c), the sister cells and cousin cells can be identified easily according to the BE values in Table 3.4: cell 1 and cell 4 are sister cells, cell 2 and cell 3 are sister cells.

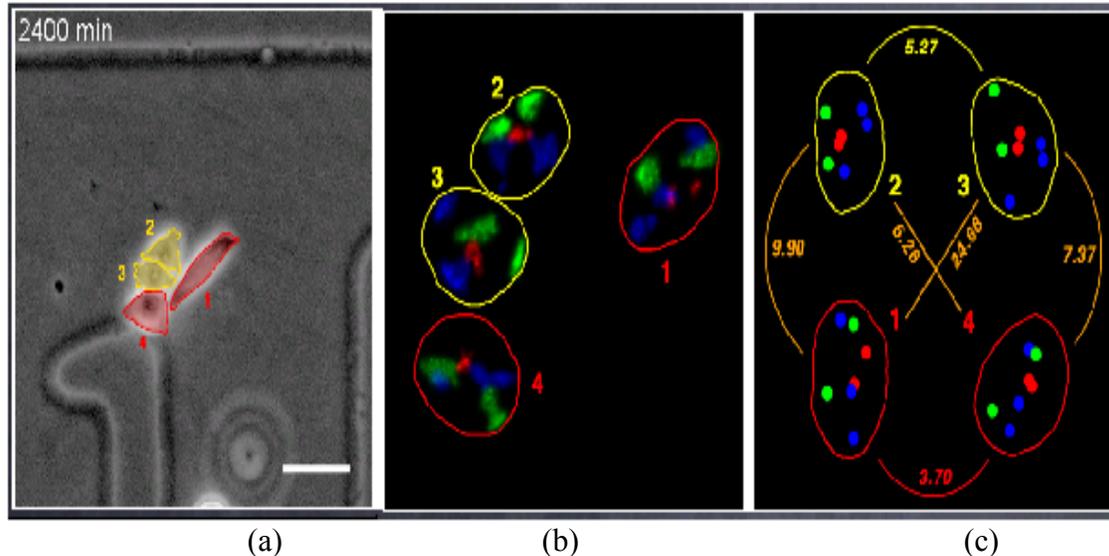


Figure 3.11 The identification of sister cells in one clone of HeLa cell line (Köhler et al. 2004 (B)). (a) Monitoring of one HeLa cell clone. (b) Projections of confocal image stacks obtained using 3D FISH technology after painting of chromosome 4 (visualized in green), 7 (visualized in blue) and 21 (visualized in red). (c) BE calculations between every two cells in this clone.

Table 3.4 The BE comparison in one 4-cell clone of HeLa cell line.

BE of every two cells in one HeLa cell clone	Bending Energy
Cell 1 and cell 2	9.90
Cell 1 and cell 3	24.68
Cell 1 and cell 4	3.70
Cell 2 and cell 3	5.27
Cell 2 and cell 4	6.26
Cell 3 and cell 4	7.37

However, in some cases, it is difficult to identify sister cell and cousin cell pairs in cell clones by Bending Energy. For example, in Table 3.5, for HMEC cell clone No. 6, among the 6 BE values, it is difficult to judge that the cell 1, 2 are sister cells or cell 1, 3 are sister cells because both pairs have smaller BE values. In this case, the judgement also depends on visualization.

The comparison of sister cells generally yields a smaller bending energy than the comparison of cousin cells. The judgement based on bending energies is more reliable compared to distance and angle measurements, although Bending Energy can not identify sister cell and cousin cell pairs for all clones absolutely. It will be interesting to test how accurately BE can identify sister and cousin cell pairs.

Table 3.5 The BE comparison in the HMEC 4-cell clone No. 6.

Combinations of every two cells in HMEC clone No. 6	Bending Energy
Cell 1 and cell 2	3.792047
Cell 1 and cell 3	7.165879
Cell 1 and cell 4	14.461073
Cell 2 and cell 3	16.229453
Cell 2 and cell 4	12.430933
Cell 3 and cell 4	20.013632

3.4.3 DD result in every cell line

In order to compare the CT arrangement of two cells in a simpler way, the difference between the distances of the geometric centers of corresponding points in every two cells was calculated. For the sake of simplicity, only the distances between homologues of the same chromosome in every cell were taken into account. In order to compare the different cell stages (Generation I-VI), the mean over all DD values inside one generation was taken as the DD of this cell stage.

3.4.3.1 HeLa cell line

The DD result of HeLa cell line is depicted in **Figure 3.12**. The DD of four-cell stage is bigger significantly than the DD of two-cell stage. This indicates that CT spatial arrangement in sister nuclei is more similar compared to cousin nuclei.

In the graphs, the dissimilarity of the arrangements increased monotonously for the different cell stages. After the 6th generation (32 cell stage), the dissimilarity of CT spatial arrangement reached a value close to RCC cells (The DD value of RCC clones is not shown here. RCC cells come from different clones and thus reflecting the dissimilarity in the whole culture of this cell line).

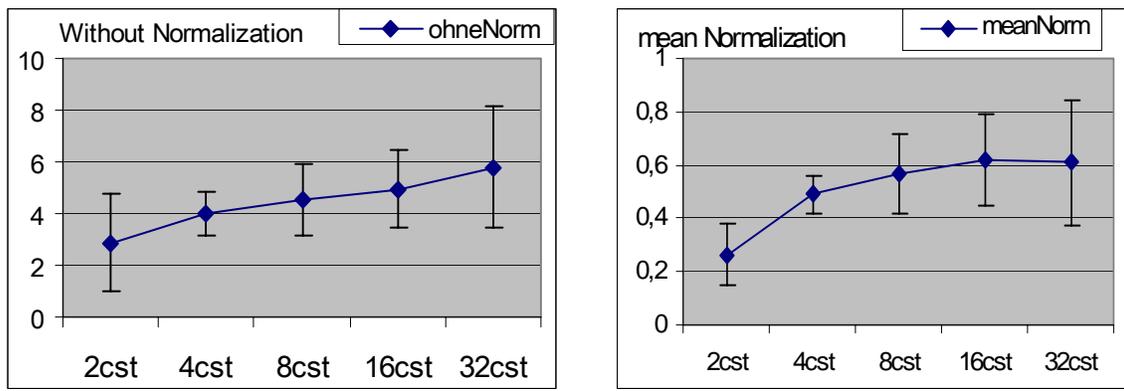


Figure 3.12 Distance Difference calculation in HeLa cell line. Left: without normalization; right: with mean Normalization.

3.4.3.2 HMEC cell line

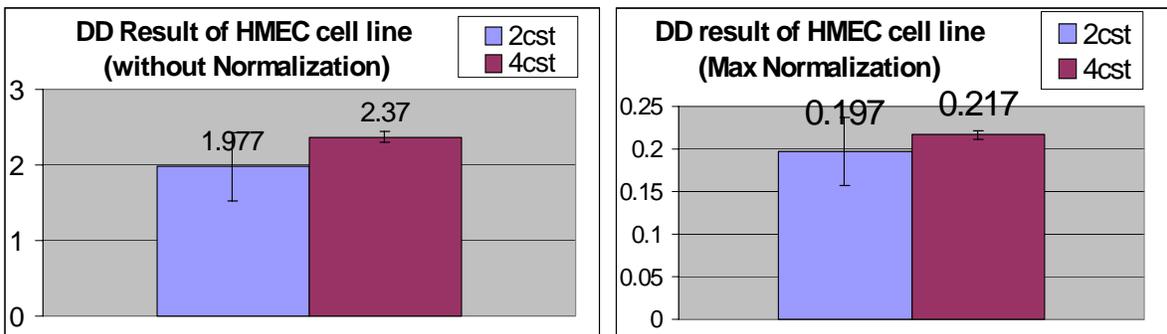


Figure 3.13 DD calculation in HMEC cell line. Left: without normalization; right: with Max Normalization.

Figure 3.13 shows the DD result of two-cell stage and four-cell stage in HMEC cell line. The DD value of four-cell stage was larger significantly than the DD of two-cell stage. This gave us the hint that the conclusion that the spatial arrangement of CTs in sister nuclei was more similar than cousin nuclei was not limited only to one cell line. The similar conclusion about the spatial arrangement of CTs in daughter cell nuclei can be drawn from different cell lines.

3.4.3.3 HFb_CB cell line

As there are only data for two-cell stage (22 cell clones) in HFb_CB cell line, it is impossible to compare the CT spatial arrangement within different cell stages. The DD calculation result with different normalizations for two-cell stage is shown in Table 3.6.

Table 3.6 The DD calculation in HFb_CB cell line.

		Without normalization	Max Normalization	PCA Normalization	Mean Normalization
HFb_CB	2cst	0.97	0.1	0.581	0.2
HFb_Ki67	2cst	2.41	0.18	no value *	0.3

* There are only 4 homologue signals in HFb_Ki67 cell nucleus.

3.4.3.4 DLD cell line

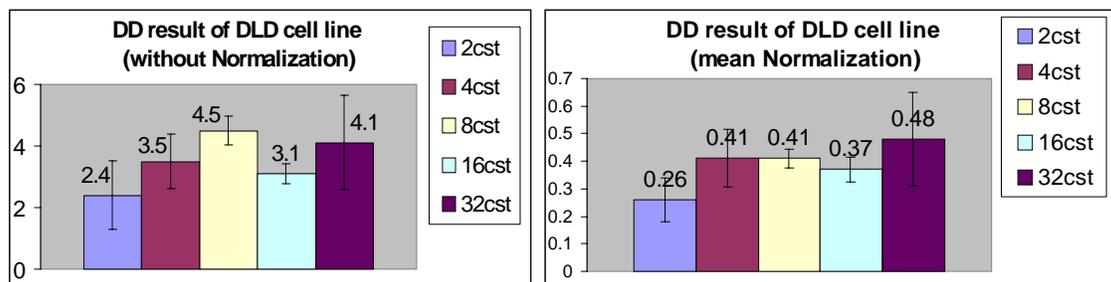


Figure 3.14 DD calculation in DLD cell line. Left: Without normalization; right: with Mean Normalization.

Another cell line to test the CT spatial configuration in daughter/cousin/grand daughter cells is DLD. In **Figure 3.14**, the average DD values in DLD clones of every cell stage

were monotonously growing up to 8-cell stage. Then the DD values waved a little bit in 16-cell stage and 32-cell stage.

3.4.4 The comparison of DD and BE approaches

From the BE result in section 3.4.1 and the DD result in section 3.4.3, the same conclusion for the different cell stages in HeLa cell line can be drawn: The spatial arrangement of CTs in sister nuclei was more similar compared to cousin nuclei. The dissimilarity of the arrangements increased monotonously and after 6 generations (32 cell stage) reached a value close to non-related cells in RCC clones. The similar conclusion can be drawn from HMEC cell line and HFb cell line.

Both approaches (DD and BE) can be used to estimate changes in CTs arrangement during clonal growth for both cancer cell lines and normal cell lines. Based on the continuous and invertible spatial transformation which maps the corresponding points in two cells on each other, BE is not influenced by nuclear shape and size difference. DD calculation depends heavily on the shape and the size of cell nucleus. Therefore, BE is a better method for the comparison of spatial CT arrangements, and DD is a necessary complement (see section 5.1.2.2).

3.5 Conclusion

The cell division cycle is under tight control (Carmo-Fonseca 2002). In order to understand the mechanism behind this tight control, new tools based on registration and non-rigid parametric motion are presented in this chapter for the quantification and visualization of chromosome order in cell nucleus of different cell lines. The conclusions drawn from the results in section 3.4 are listed here:

- (1) CT arrangements are more similar in sister cells than in cousin cells.
- (2) In HeLa cell clones, the dissimilarities among cells in every clone accumulate with every new cell division. This dissimilarity in 32-cell clones reaches the level of dissimilarity in randomly chosen cells.

A judgement concerning the similarity of spatial arrangements of CTs in cell nuclei of different cells is difficult considering the local and global variations and irregularities of nuclear shape, as well as of the shapes and positions of individual CTs. However, the application presented in this chapter shows that spatial registration can be used in the study of chromosome arrangement during different cell divisions. For a quantitative, statistically validated analysis we defined and compared two measures based on distance difference and on bending energy. This allowed us to compare the different degrees of preservation of the spatial arrangement of chromosomes in the cell nucleus of different generations. BE method and DD method are sensitive enough to quantify the huge amount of data which describe the chromosome distributions of different cell generations. The use of the geometric centers of CTs appears as a reasonable starting point for a quantitative description of the similarities and dissimilarities of CT arrangements in different cell nuclei. The presented data show a change of chromosome arrangement that accumulate during clone growth, in particular with every new mitotic division.

As the first application of BE method into the study of chromosome spatial distribution during different cell divisions, it provides a new approach for the further research related to chromosome order inheritance and cell nuclear dynamics.

Chapter 4 Quantitative analysis of the spatial distribution of gene *MLL* and its translocation partners

4.1 The *MLL* gene and some of its translocation partners

Chromosomes in human tumor cells are often abnormal. The specific recurring chromosome aberrations, such as translocations, are often associated with a particular type of leukaemia, lymphoma or sarcoma (Rowley 2001). As more of the genes identified at translocation breakpoints were found to be oncogenes, a lot of work have been done to identify, define and analyze the genes involved in translocations (for example Carlo Croce 1982, Miyoshi et al. 199, Rowley 1998, Knudson 2001, Elliott et al. 2002, Mitelman 2000).

4.1.1 The introduction to gene *MLL*

One of these important genes involved in translocations is *MLL* (mixed lineage leukemia) on chromosome 11 band q23. Chromosome translocations involving band 11q23 are clinically characteristic in the sense that they are seen particularly in infants and are sometimes biphenotypic, indicating the presence of lymphoid and myeloid elements (Hudson et al. 1991, Cheng et al. 2001). The *MLL* gene (Ziemin-Van Der Poel et al. 1991) is involved in chromosome translocations in ~15% of patients with acute myelogenous leukemia (AML) and acute lymphocytic leukemia (ALL) (Strissel et al. 2000).

MLL gene spans ~120 kb (Wiedeman et al. 1999, Strick et al. 2000, Nilson et al. 1996). All translocation breakpoints within the gene *MLL* occur in an 8.3 kb fragment called the breakpoint cluster region (BCR) (Thirman et al. 1993, Rowley 1998). The generation of a chimeric transcript consisting of 5' *MLL* and 3' sequences of the gene on the partner chromosome seems to be the critical feature of these chromosomal rearrangements (Luo et al. 2001).

4.1.2 Some translocation partners of gene *MLL* (*AF9*, *AF4*, *AF6*, *ENL*, *ELL*)

In general, *MLL* translocations are the result of an illegitimate recombination process leading to reciprocal fusions of unrelated translocation partner genes. Over 60 different *MLL* gene translocations have been cytogenetically reported (Silvana et al. 2004, Debernardi et al. 2004), and *MLL* gene translocations are assumed to be the initial step of the malignant transformation of hematopoietic precursor cells leading either to acute myeloid or lymphoblastic leukemia (Bursen et al. 2004).

The positions of five *MLL* translocation partners were analyzed in this thesis. The location of gene *MLL* and some of its translocation partners are shown in **Figure 4.1**.

The human *AF9* (= *ALL-1* fused gene on chromosome 9) gene (Iida et al. 1993) at 9p22 is one of the most common fusion partner genes with the *MLL* gene, resulting in the t(9;11)(p22;q23). The *AF9* gene is >100 kb and two BCRs have been identified.

The *MLL-*AF9** fusion gene, which associated with AML as well as therapy related AML (t-AML), rarely with ALL, plays a critical role in stem cell development and leukemogenesis (Corral et al. 1996, Dobson et al. 1999, Joh et al. 1999).

The reciprocal translocation t(4;11)(q21;q23) is one of the most frequent *MLL* translocations known today (Bursen et al. 2004) and is recurrently found in high-risk infant acute lymphoblastic leukemia and early childhood. This translocation t(4;11) involves the human *MLL* and *AF4* (= *ALL-1* fused gene on chromosome 4) genes which covers a 300 kb genomic region on chromosome 4 band q21 (Gu et al. 1992, Bursen et al. 2004).

The translocation t(6;11)(q27;23) is one of the most common translocations observed in patients with AML, or chronic eosinophilic leukemia (Ann et al. 2001), infant acute monocytic leukemia (Akao et al. 2000) and other leukaemias. The *AF6* gene (Prasad et al. 1993) located on chromosome 6q27, is the fusion partner of the *MLL* gene in this translocation (Tanabe et al. 1996). It is possible that the *AF6* protein can be involved in signal transduction at special cell-cell junctions (Prasad et al. 1993) and *MLL/AF6* fusion protein is perhaps associated with cell proliferation (Joh et al. 1997).

The breakpoint in chromosome 19p can locate either at 19p13.1 or 19p13.3, thus could involve either of two genes: *ELL* (= 11–19 lysine-rich leukemia gene) on 19p13.1 or *ENL* (= 11–19 leukemia gene) on 19p13.3 (Cheng et al. 2001).

The (11;19)(q23p13.3) translocation, associated with de novo t-AML, juxtaposes the 5' sequences of the *MLL* gene to the 3' sequences of the *ELL* gene, forms the *MLL-ELL* chimeric gene, and results in the formation of an in-frame MLL-ELL fusion protein (Thirman et al. 1994). As a nuclear protein, ELL protein is an RNA polymerase II elongation factor that has been implicated in oncogenesis (Shilatifard et al. 1996).

Gene *ENL*, one common TP of gene *MLL* involving ALL and rarely, AML (Tkachuk et al. 1992, Rubnitz et al. 1994), encodes a protein with transcriptional transactivation properties. Together with the other two more common translocations MLL-AF9 [t(9;11)] and MLL-AF4 [t(4;11)], MLL-ENL fusion protein is one of the three most frequently found fusions in leukemia cases.

In this chapter, the quantitative analysis of the spatial distribution of five of *MLL* translocation partners, *AF4*, *AF6*, *AF9* (9p22), *ENL*, *ELL*, and a set of control loci (2q33, 2q35, 7q22, and 8q34) which were analyzed using 3D FISH probes were presented.

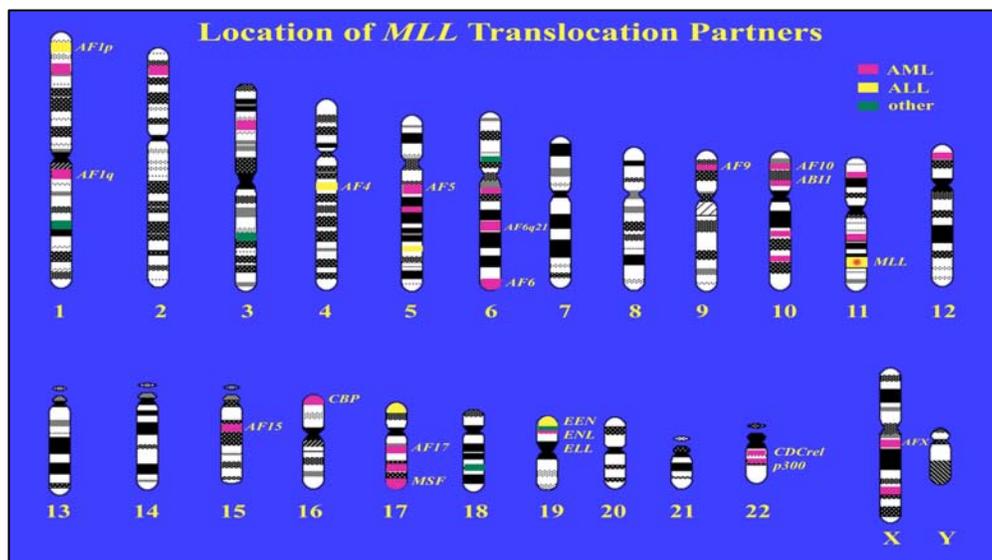


Figure 4.1 The location of gene *MLL* and its translocation partners (Murmam 2004).

4.2 Data description

4.2.1 Cell culture, cell cloning, 3D-FISH and surface reconstruction

Cell lines (see Table 4.1) were seeded on gridded coverslips, monitored throughout the clonal growth, and finally fixed. 2D-FISH and 3D-FISH were performed with painting probes specific for gene *MLL*, its five translocation partners (*AF4*, *AF6*, *AF9*, *ENL* and *ELL*) and four control loci (2q33, 2q35, 7q22, and 8q34). The control loci were chosen from chromosome regions that have not been reported to translocate with gene *MLL* or 11q23. Confocal image stacks for all three-color channels were recorded with a confocal microscope (SP2 AOBS, Leica) and software LCS 25v1347. Median filtering was applied to reduce noise in the raw images and the 3D-reconstructions of confocal image stacks were performed using Amira™ 3.1 (TGS).

4.2.2 Cell lines used in this study

Table 4.1 Cell lines used in this chapter to calculate the gene spatial distribution.

cell line	description	cell line	description
990J1	Diploid	HDF	Diploid
Triggs	Diploid	PB (peripheral blood)	Diploid
		CCL-157 (male Indian muntjac)	Diploid
KG-1	Hypodiploid	FIM (female Indian muntjac)	Diploid
KG-1A	Diploid	PCF (female Chinese muntjac)	Diploid
U937	Hypotriploid	CCL44 (male cattle)	Diploid
Mono Mac6	Hypotetraploid t(9;11)	Nalm6	Diploid
THP-1	t(9;11)	SKW 6.4	
Rs4;11	t(4;11)	Jurkat, clone E6-1	Pseudodiploid
MV411	t(9;11)	HUT-78 subclone H9	triploid
ML-2	t(6;11)		

To address the question whether a difference in chromosomal numbers has an effect on the localization of genes and whether the conclusions drawn in this chapter could be generalized to different mammalian species, three different species were chosen: two relatively closely related species of deer, *Muntiacus muntjak* and *Muntiacus reevesi*, with

a great difference in chromosome number, and the well characterized human species. *M. reevesi* serves as a control for a species with the same chromosome number as the human. Table 4.1 shows a list of all cell lines in these three different species used in this chapter. In every cell line, there are at least five signals, for example, two signals for gene *MLL*, two signals for its translocation partner, and one signal for the whole cell nucleus.

4.3 Result and conclusion

4.3.1 The distance between the nuclear center (NC) and genes

Fibroblast nuclei are oval shaped and rather flat, so it is difficult to analyze the locations of genes using shell-analysis methods. Therefore, to describe the position of a gene within a fibroblast, the distances of the genes to the geometric center of the nucleus (also called nuclear center, NC) were determined. From the calculation result we can see, that *MLL* had the largest median distance to the NC, whereas *AF4* and *AF9* share a similar shorter distance. This distribution pattern is not the same as in Triggs cells, which are hematopoietic cells.

4.3.2 Distance among genes in hematopoietic cells

To describe the position of two genes relative to each other, the distances of gene *MLL* to its various translocation partners were measured, and the distance results were analyzed. In conclusion it appears that the distance of the investigated genes is a function of the relative distribution pattern of a gene in the nucleus, but not the preferred location of the genes next to each other. The actual closeness of 2 μm or less between *MLL* and a translocation partner was a rare event.

4.3.3 The shortest distance between genes and nuclear surface

Chromosome-specific FISH paints lack a fixed reference point, which makes the analysis of chromatin/gene location difficult. In order to determine the location of the genes/loci, the shortest distance between genes to the nuclear surface was calculated.

Table 4.2 The statistical test of shortest distances between one gene and the nuclear surface.

Cell line	α value of Wilcoxon Signed Rank test		
	AF4-AF9	AF4-MLL	AF9-MLL
MMV (= female Indian muntjac fibroblast)	0.00097	0.17762	0.00000
MRE (= male Chinese muntjac fibroblast)	0.14696	0.00137	0.05869
Human cell line KG1			0.00499
Human cell line KG1A			0.02741

Figure 4.2 shows the positions of the gene *MLL* (green) and *AF9* (red) signal relative to the nuclear surface in Triggs cells. In Figure 4.2 (a), the gene *AF9* (red) has a higher probability to be closer to the nuclear surface than the *MLL* gene. The shortest distances of gene *MLL* and *AF9* are significantly different from each other in four different cell lines, using the statistical method "Wilcoxon Signed Rank test" (Hollander et al. 1973) (see Table 4.2). The null hypothesis (H_0) for this test is that there is no significant difference between the shortest distances of every two genes. In Table 4.2, the α values for the comparison of the shortest distances of gene *MLL* and gene *AF9* in most of the cell lines are always less than 0.05.

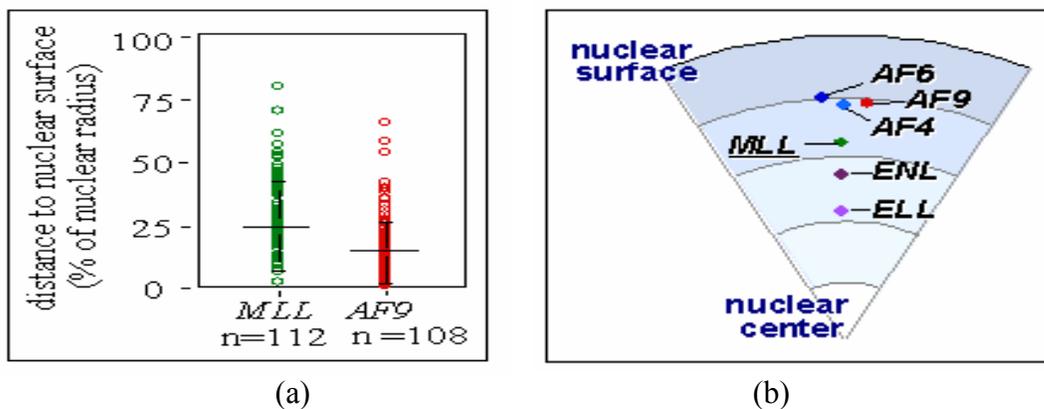


Figure 4.2 The shortest distance between genes and nuclear surface in Triggs cells. (a) The shortest distances of gene *MLL* and *AF9* are significantly different. (b) The 3D

localizations of gene *MLL* and its translocation partners show a characteristic distribution pattern in cell nuclei.

Statistical confidence (p) values were obtained from the shortest distance calculation using Wilcoxon Signed Rank test. The α value less than 0.05 indicates that there is significant difference between the shortest distances of two genes.

In Figure 4.2 (b), the mean positions of the genes were plotted with respect to the position of *MLL* in each experiment. It is shown that the distribution of the three most common translocation partners *AF6*, *AF9* and *AF4* is remarkably similar. Another two translocation partners also have their characteristic distance from the nuclear surface.

From this figure we can draw the conclusion that the *MLL* gene and its translocation partners show characteristic distances to the nuclear surface in fixed hematopoietic cells, and the localization of gene *MLL* seems to follow a similar distribution pattern in all analyzed cells. This conclusion is not limited only to one cell line. For the different cell lines, we can get the same conclusion.

4.3.4 Angles among genes

In order to investigate the spatial relationship between the genes and the nuclear center, in addition to the distance methods mentioned above, the angle calculations were used. The angles between two genes around the nuclear center are in $[0, 180)$. Statistical method "Kolmogorov-Smirnov test" (Kolmogorov 1933, Smirnov 1936) was employed in four different cell lines (MMV, MRE, KG1, KG1-A), it was found that the angles among genes were randomly distributed in $[0, 180)$.

From the results and the evaluations described above (for the details and graphs of some calculation results, see (Murmam 2004)), we draw the conclusion from the translocation data:

1. In hematopoietic cells, the gene *AF9* is in general more closely positioned to the nuclear surface than the *MLL* gene. Despite the cell materials' differences in maturation state, cell lineage, and chromosome number, the localization of each gene and

chromosomal locus showed a characteristic distribution pattern in the interphase nucleus in all studied hematopoietic cells.

2. Spatial arrangement of translocation partner genes within the interphase nucleus is not random and has specific patterns, and these distribution patterns are not limited to one cell line. The gene positions might be specific for the respective cell type.

Chapter 5 Discussion and future work

5.1 Discussion of the proposed computational tools

5.1.1 Geometric center and the center of gravity

5.1.1.1 Is the geometric center method a crude method?

In this thesis, the decision was made to analyze the positions of chromosomes or genes in form of their geometric centers instead of their surface points. A mathematical analysis based purely on the geometric centers ignores the spatial extensions of the considered objects. For example, there are maybe ten thousands of points at the reconstructed surface of chromosome, but all points are replaced by just one point which is the geometric center of this chromosome, and the details of the spatial extension of chromosome are totally neglected. Therefore, the following questions arise: Why is the geometric center used instead of the whole chromosome volume? Is the geometric center a crude method to quantify the spatial distribution of chromosomes or genes?

For a detailed comparison of topological difference (measured based on relative positional changes of the location of geometric centers) of shape variations of chromosome territories, it is necessary to establish more accurate tools taking into account the spatial extension of the chromosome territories. However, a judgement concerning the similarity of spatial arrangements of chromosome territories in different cell nuclei is difficult considering the local and global variations and irregularities of nuclear shape, as well as of the shapes and positions of individual chromosome territories.

In order to focus on the distribution of chromosome territories during cell divisions at the genome level, it is necessary to pay more attention to the whole topology of chromosome distribution, but not every detail of chromosome spatial extension. Therefore, the shape differences of chromosome territories are neglected, and the geometric centers of the homologues are used instead of their surface points. The geometric centers are used in order to detect the global spatial distribution of chromosomes at the genome level.

Another reason to use geometric centers is that geometric centers simplify the quantitative evaluation as only one point is used instead of thousands of points. Although the geometric center can not provide enough spatial information of the entity studied, and the shape of chromosomes or genes can vary significantly even when their geometric centers are identical (although it never happens), geometric centers keep the topology of the spatial entities distribution.

5.1.1.2 The geometric center is different from the center of gravity

Given a system of material points $M_i (x_i, y_i, z_i)$ with the masses $m_i (i=1, 2, \dots, n)$, center of gravity is defined as (Harris et al. 1998):

$$x = \frac{\sum m_i x_i}{\sum m_i}, y = \frac{\sum m_i y_i}{\sum m_i}, z = \frac{\sum m_i z_i}{\sum m_i}.$$

In the field of image processing, the center of gravity in one image is defined as:

$$x = \frac{\sum p_i x_i}{\sum p_i}, y = \frac{\sum p_i y_i}{\sum p_i}, z = \frac{\sum p_i z_i}{\sum p_i}.$$

$p_i (i=1, 2, \dots, n)$ is the pixel intensity of the pixel at the position x_i .

From the formulars it can be seen that geometric center and the center of gravity are different, but the geometric center can be thought of as the center of mass of an object with an equal intensity of 1.

5.1.2 Advantages of these tools

In order to define and evaluate the exact 3D positions and distribution of genes or chromosomes, it is necessary to develop new computational tools to quantify 3D position, distances, angles, and dynamic parameters for chromatin distribution data sets. At the same time, in order to quantify the similarity or difference of chromosome spatial distribution within interphase nuclei, it is necessary to develop new quantitative tools which can estimate the transformation that maps spatial points in different volume based on the motion model of thin-plate spline transformation. The new quantitative tools should allow the quantification of similarity of chromosome spatial distribution in different cell nuclei or during different cell divisions. These tools should be able to define

and evaluate precisely the spatial arrangement of potential translocation partner genes within the interphase nucleus.

For the first time, by applying point set registration and the bending energy, the inheritance of chromosome spatial distribution could be revealed during several cell divisions in different cell lines. Therefore the methods presented in this thesis are a novel application of registration methods for biological data sets. It is also the first time that the spatial distribution of genes, with focus on gene *MLL* and some of its translocation partners, could be analysed and quantified within the 3D space of interphase nuclei in hematopoietic cells. These new tools (BE, DD and even distance calculation) also serve as powerful tools to check the errors in the original data.

The tools described in this thesis are valuable for a quantitative analysis of spatial distribution of cell nuclear structure. An essential feature of these computational tools is the capability to calculate, quantify and visualize the distances and angles of 3D reconstructed entities based on confocal image stacks. The signals in the image stacks can be genes, replication loci, chromosomes, proteins or other entities which have biological meaning. One application of these computational tools introduced is to describe exactly the positions, distances and angles among gene.

5.1.2.1 Use of the Bending Energy is one possibility to evaluate the similarity of chromosome spatial distribution during different cell divisions

In cell biology, parametric registration and thin-plate spline transformation have been used mainly for automated correction of rotational and translational movements in a time series. But in this thesis, as a parameter reflecting a degree of space transformation is required for the best match of two chromosome territories, the bending energy is applied for the evaluation of chromosomal spatial arrangements during cell divisions.

The use of the geometric centers of chromosome territories appears as a reasonable starting point for a quantitative description of the similarities/dissimilarities of chromosome territory arrangements in different cell nuclei.

For a quantitative, statistically validated analysis two measures based on distance differences and on the bending energy were used. The bending energy overcomes the problem that nuclear shape and size can vary largely in different cells, when defining a

procedure to assess quantitatively the similarity of spatial CT arrangements in two cell nuclei. The compensation for a global deformation and/or a measure to distinguish local variations in the spatial configuration from global deformations is needed. Therefore BE is a measure well-suited to compare chromosomal arrangements for which it is applied for the first time in this thesis.

Distance difference calculation depends heavily on the shape and the size of the cell nucleus. But if the shape and the size of the cell nucleus do not change a lot in one cell line, distance difference calculation can serve as a good tool to evaluate the similarity of chromosome territory distribution in different cell nuclei. It is also a tool for data checking as described in the next section.

These two methods (or two computational tools: Bending Energy and distance difference) allowed us to compare the different degrees of preservation of the spatial arrangement of chromosomes in the cell nucleus for different generations of cells. The presented data and result in Chapter 3 show a change of chromosome arrangement that accumulate with every new mitotic division.

In the next step, it is possible to envisage to sample points on unit spheres around each geometric center as well as on CT surfaces.

5.1.2.2 Error checking in the original data

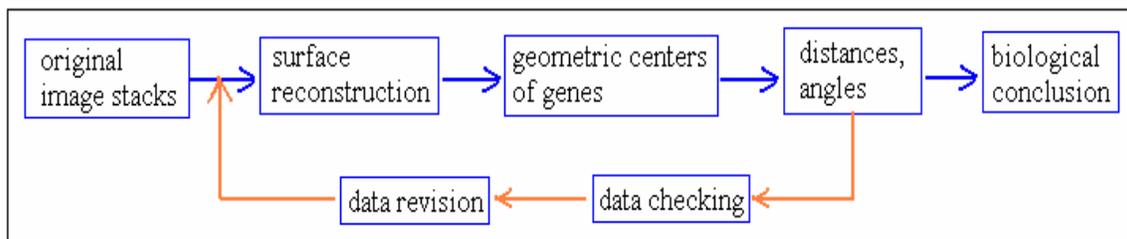


Figure 5.1 Relationship between data checking (red arrows and red lines) and computational tools developed in this thesis

The accuracy of the data is important for the further evaluation, and the accuracy is especially important to help biologists to draw the correct final conclusion. However, as the threshold for threshold-based segmentation was selected manually by researchers, it is easier to make

some mistakes from the segmentation and 3D surface reconstruction, compared to the automatic segmentation.

The computational tools developed in this thesis are useful to detect errors made in manual segmentation of the data and to improve the precision of the original data. For example, for chromosomal position distribution calculation, the normal bending energy (BE) values involved is always less than 50. If the BE value is bigger than 100, there is probably an error in the original data. If one value of distance difference is more than 20 μm (20 μm is even larger than the diameter of whole cell nucleus), obviously there is an error in the original data or in the manual segmentation.

The bending energy alone is not enough to check all errors in the data. The distance difference (DD) also serves as a necessary tool to check the errors. Sometimes the following case can happen: The BE value is within the range, for example, less than 10, but the DD is pretty huge, more than 50 μm , this indicates that there is an error in the data. So the DD calculation is a necessary complement for data checking.

For the spatial distribution evaluation of different genes such as the *MLL* gene and its translocation partners, there is also the possibility to check this data with these tools. For example, if the shortest distance between the nuclear center and nuclear periphery is more than 20 μm , which is even larger than the diameter of cell nucleus, there should be an error in the original data. If the distance between gene A and gene B is identical to the distance between gene A and gene C, there should be a mistake in the 3D surface reconstruction of gene B or gene C ---- gene B and C are same gene, but they are saved using two different names because of the mistakes made by user manually.

After the error checking, the data can be corrected manually. The proposed tools (BE, DD and distance calculation) are able to check errors in the original data. By the use of DD, BE and distance values, the precision of the revised data was improved obviously (**Figure 5.1**).

5.1.3 Improvements of these tools

5.1.3.1 Triangulation: Marching cube algorithm

To triangulate the 3D surface of chromosomes/genes, the algorithm used for 3D surface reconstruction is the marching cube algorithm (Cline et al. 1988), which is the most

commonly used method. In this algorithm, the 3D biological structure is defined by a threshold throughout the data set in order to construct an isosurface. This 3D surface reconstruction approximates a selected structure by a list of triangles or other polygons. By changing the viewing direction interactively, the user can check the displayed 3D biological structure within a graphic interface on the computer.

Rendering algorithms were well developed, but the generation of a polygon list which represents the surface in an appropriate way can still be difficult. One reason is that the surface of many biological structures (like replication domains, chromatins) cannot be defined using a single intensity value. For the 3D surface reconstruction of some biological data, the marching cube algorithm results in the loss of useful information (Gerlich et al. 2003 (B)). So for some specific biological data (for example, the detailed spatial extension of genes) which can not be triangulated using one single intensity threshold, the algorithm for surface reconstruction should be improved.

5.1.3.2 Segmentation

The prerequisite to use the computational tools presented in this thesis is to get the 3D surface reconstruction from confocal image stacks. Thresholding is suitable to segment objects from a background with different gray value intensity compared to the objects. In our work the threshold for threshold-based segmentation was selected manually by the user. The segmentation of genes or chromosome homologues was finished by hand, according to the user's experience. An experienced user can use manual operation to judge how to do segmentation in a biologically meaningful way. This is especially useful when the relevant signals in the image stacks are too weak, or there are some extra signals that mix together with the signals that have biological meaning.

However, sometimes user's manual operation can lead to the following errors:

- a. In the case of dealing with a huge amount of data, manual operation is time consuming. User can make mistakes when surface construction, data naming and data saving are done manually.

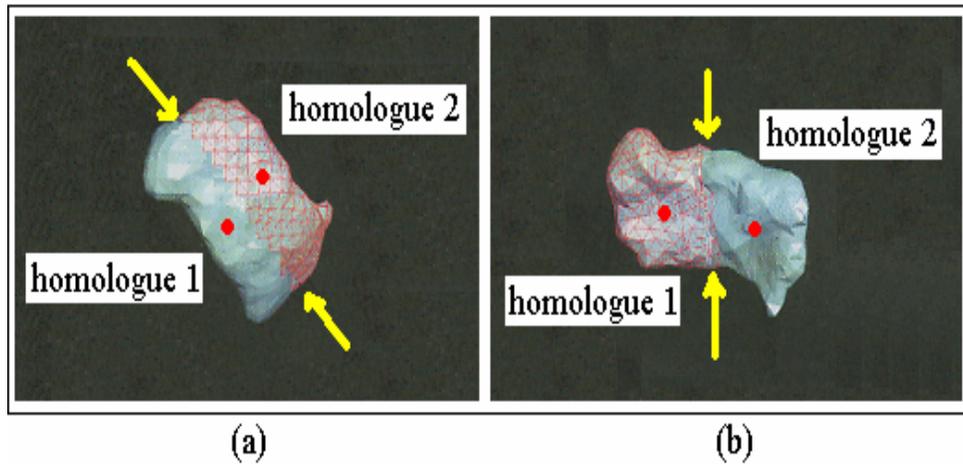


Figure 5.2 Two different ways to do segmentation for the two homologues in one chromosome. Yellow arrows in (a) and (b) show the directions for segmentation. Two different ways lead to different 3D positions of homologue geometric centers (red points) in (a) and (b).

b. In order to obtain the 3D surface reconstruction of two homologues in one chromosome, it is necessary to divide one chromosome into two different parts (see the data description in Chapter 3). If the two homologues are far away from each other, the segmentation will be easier. However, sometimes it is difficult to do the segmentation. For example, in **Figure 5.2** (a), the two homologues (the part in blue and the part with red lattice) in one chromosome are so close to each other that it is almost impossible for a user to distinguish these two homologues.

If these two homologues juxtapose to each other, should they be arranged as shown in Figure 5.2 (a), or Figure 5.2 (b)? Should the segmentation be done horizontally as shown in Figure 5.2 (a), or perpendicularly as shown in Figure 5.2 (b), or in some other way? If these two homologues do not juxtapose to each other but twist together, how do they twist? Obviously, the misjudging of two homologues here will lead to the wrong segmentation. After segmentation, the geometric center (red point) of homologue 1 in (a) changed a lot, compared to the geometric center of homologue 1 in (b). Similarly, the geometric center of homologue 2 changed as well. So in the case of Figure 5.2, it is difficult to do the segmentation in a correct way. Intelligent automatic segmentation tools which can quantify such complicated biological data are necessary.

c. Even in the case that the user knows how to divide a chromosome into two parts (e.g., the user does know that the chromosome should be divided along the direction shown by yellow arrows in Figure 5.2 (a)), it is still difficult to do the segmentation manually in Amira 3D viewer in an accurate way: Amira does not provide the function that the user can do the segmentation manually with the interactive changing viewing direction. This will also bring some tiny errors into the segmented surfaces.

Thresholding is a simple and often very efficient way to segment an image. But as mentioned above, a single intensity value can not define the surface of some biological structures. If threshold-based segmentation is suboptimal, it could be improved by an active contour (snake method) which refines the boundary initially assigned to an object by integrating more global image features. Recent efforts have been made to make automatic parameter adjustment available for snake algorithm (e.g. Gebhard et al. 2001, Gebhard et al. 2002, Gerlich et al. 2003) as snake algorithm depends strongly on the precise parameter settings determined by specific applications, especially complex biological data sets.

5.2 Models for chromosome positioning during interphase and mitosis

Since the first model of chromosome position was developed by Rabl (Rabl 1885), a lot of quantitative models have been developed thereafter.

5.2.1 Models for interphase chromosome positioning

(1) Rabl configuration

The territorial organization of chromosomes in interphase cells was originally proposed by Rabl and Boveri more than a century ago (Rabl 1885, Boveri 1909). In 1885, based on his observations of salamander cell division, Carl Rabl proposed that the centromere–telomere orientation of chromosomes observed during anaphase is maintained throughout the cell cycle. As a consequence, in the interphase cell centromeres and telomeres are at opposite sides of the nucleus (**Figure 5.3**). The Rabl configuration is found in *Drosophila*

and plants (Hochstrasser et al. 1986, Abranches et al. 1998). In mammalian cells, the Rab1 configurations are relatively rare during interphase. Yeast typically has a Rab1-like configuration of clustered centromeres and clustered telomeres (Jin et al. 2000).

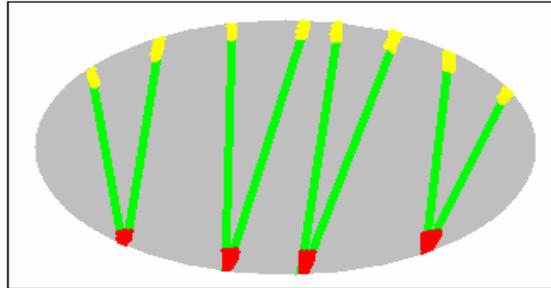


Figure 5.3 Rab1 configuration: centromere (red) and telomeres (yellow) of every chromosome (green) are aligned at opposite sides of the nucleus.

The Rab1 configuration is established during anaphase and is maintained during interphase. It ensures that the orientation of chromosomes within a nucleus is preserved, which probably aids the maintenance of chromosomal integrity. The specific positioning of chromosomes has been suggested to have functional consequences by facilitating the alignment of homologues during meiosis (Zickler et al. 1999).

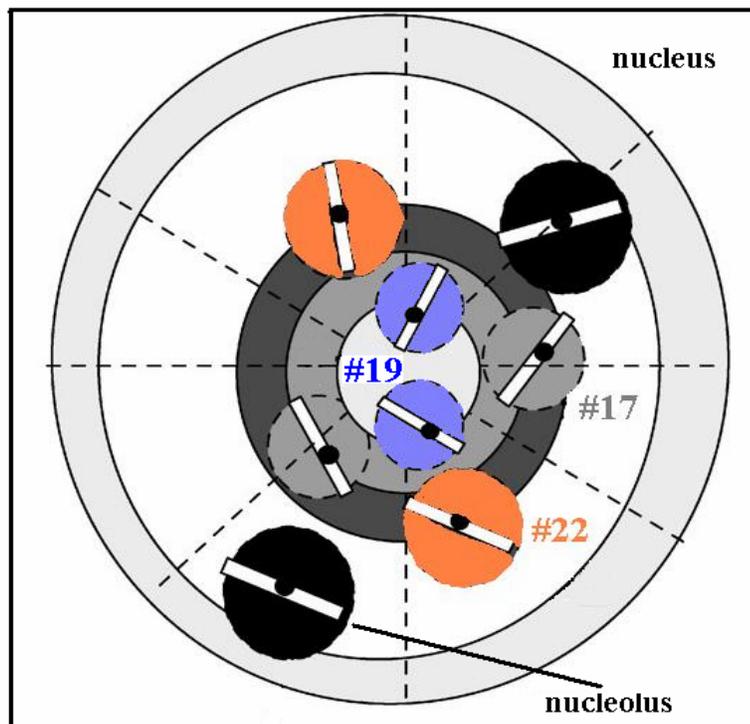


Figure 5.4 Radial positioning of chromosomes. Chromosome territories distribute towards either the periphery or the interior of the cell nucleus according to their size or gene density. Adapted from Kreth et al. 2004.

However, the alignment of chromosomes in a Rabl configuration does not appear to be important for the control of gene expression. Active genes on chromosomes arranged in a linear Rabl configuration in *Drosophila* embryonic nuclei are found along the entire chromosome axis and there is no correlation between their expression level and their position along the chromosome (Wilkie et al. 1999).

(2) Radial positioning

Within human and primate lymphocyte cell nuclei, from the center to the nuclear envelope, the gene rich chromosomes are preferentially found in the nuclear interior, and gene poor chromosomes generally localize closer to the nuclear envelope (**Figure 5.4**). This model is not only specific to HSA 18 and 19 (Croft et al. 1999, Cremer et al. 2001), but also holds for all human chromosomes (Figure 5.4) (Boyle et al. 2001). The radial positioning of chromosomes in the cell nuclear volume suggests a relationship between the gene density of one CT and its distance to nuclear center. But radial positioning has only been documented in a few cell types and there seems to be cell-type-specific differences in CT positioning (Cremer et al. 2001). Based on the radial positioning model, a "spherical 1 Mbp chromatin domain" model was presented to relate chromosome positioning and gene density in a quantitative way (Kreth et al. 2004).

(3) Relative positioning

In this model (**Figure 5.5**), chromosomes occupy preferential positions within the nucleus and, as a consequence, relative to each other. Evidence for relative positioning comes from observations of mitotic rosettes (or mitotic ring) (Nagele et al. 1995, Allison et al. 1999). During metaphase, chromosomes were arranged in a ring in the metaphase plate. Homologues often occupy diametrically opposed positions in the ring. These observations were extended into interphase cells, in which the homologues of human chromosomes 7, 8 and 16 were found in diametrically located positions within the interphase nucleus of quiescent cells (Nagele et al. 1999). However, analysis of the

relative positioning of chromosome territories in chicken cells revealed no preferential patterns (Habermann et al. 2001).

None of these three models is mutually exclusive. For example, methods to detect radial positioning do not generally probe for relative positioning and vice versa (Fig. 5.4 and 5.5) (Parada et al. 2002 (B)).

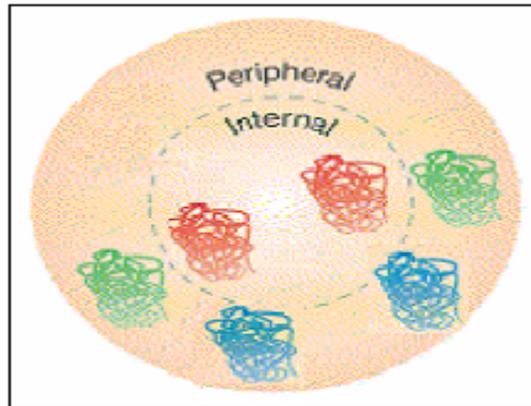


Figure 5.5 Relative positioning: chromosome territories occupy the preferential positions relative to each other. Adapted from Parada et al. 2002 (B).

5.2.2 Models for mitotic chromosome positioning

During mitosis of many higher eukaryotes, the spatial order of the nucleus is lost when the genome is packaged into metaphase chromosomes, and this order is reconstituted in daughter cells after cell division. The following five questions arise from the genome level:

1. In prophase, does the position change during chromatin condensation?
2. In metaphase, is there any position changing when condensed chromosomes are arranged to form the metaphase plate?
3. In anaphase, is there any position changing when sister chromatids move individually towards opposite poles of the spindle to form daughter nuclei?
4. In early G1 phase, is there any position changing when chromosomes decondense in daughter cells?
5. Is the spatial arrangement of chromosomes reversibly broken down and reconstituted in daughter cells after cell division?

In order to answer these questions, different models have been proposed, see **Figure 5.6**.

Model 1: Physical linkage between chromosomes

Chromosome positioning in metaphase rosettes of human cells is not random (Nagele et al. 1995). Chromosomes are physically attached to each other during metaphase (Maniotis et al. 1997), and this chromosome physical linkage in interphase and mitosis contribute to the conservation of neighborhood relations. Symmetrical chromosome positions in sister cells indicate that chromosome position reconstruction depends on the metaphase configuration (Sun et al. 1999). During chromatin condensation, relative neighborhoods are maintained (Gerlich et al. 2003 (A), Chaly et al. 1988). During chromatin decondensation, there are no major relative positional changes (Manders et al. 1999, Manders et al. 2003).

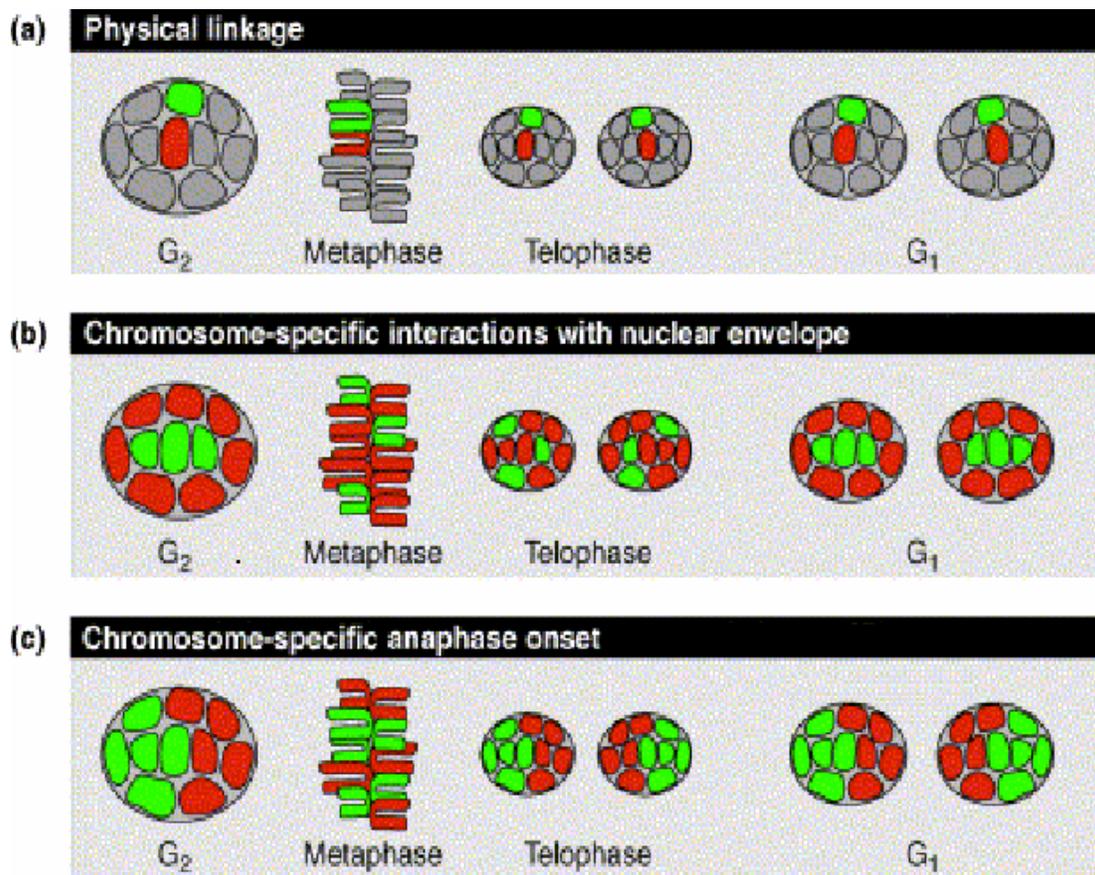


Figure 5.6 Three models for mitotic chromosome positioning (Gerlich et al. 2003 (C)). (a) Physical linkage. (b) Chromosome-specific interactions with nuclear envelope. (c) Chromosome-specific anaphase onset.

This model could explain how chromosomal neighborhoods could be preserved throughout mitosis. However, one study could not confirm a non-random metaphase rosette configuration (Allison et al. 1999).

Model 2: Chromosome tethering by the nuclear envelope

During mitosis, chromosome tethering with broken down nuclear envelope makes the chromosome positioning random. In early G1, by specific interactions of some chromosomes with the reforming nuclear envelope (gene-poor chromosomes tether to peripheral positions), chromosomes establish radial positioning. There are chromosome repositioning movements when the nuclear envelope reforms during early G1 phase of live cells (Walter et al. 2003). After cells had passed through a mitotic division, the typical peripheral localization of gene-poor chromosomes was observed (Bridger et al. 2000). This model could explain how a radial order is established during mitosis, but it can not explain why chromosome arrangements in clonally descendent cells resemble each other more than in unrelated cells.

Model 3: Chromosome-specific timing of segregation

Along the spindle axis, during chromosome congression to the flat metaphase plate, the spatial order information of chromosomes is lost, but this order (non-random relative chromosome positions) is re-established during segregation in early anaphase by chromosome-specific anaphase onset, such that the order in daughter nuclei were again similar to the mother nucleus. Chromosome-specific timing of segregation could determine their positions as the initiation of poleward chromosome movements correlates with the position along the spindle axis in the daughter cell (Vig 1981, Gerlich et al. 2003 (A)). The increasing amounts of pericentromeric heterochromatin could delay the timing of chromosome segregation (Tanaka et al. 1999).

Along the metaphase plate, positions are maintained in daughter nuclei by essentially linear congression and segregation. This model predicts that interphase positions mainly depend on centromeric composition rather than on chromosomal arm sequences.

This model is consistent with both modes of radial positioning and preferred neighborhoods. This model does not exclude the possibility of chromosome-specific

interactions with the nuclear envelope, which might further modify postmitotic chromosome positioning during early G1.

5.2.3 A new model for chromosome position inheritance during cell divisions

A new model of chromosome position inheritance during cell divisions is shown in **Figure 5.7**. According to Gerlich et al. (Gerlich et al. 2003 (A)), the dissimilarity of chromosome arrangement after several cell divisions keeps stable. This can be represented by the purple line 2 in this figure. Another possibility for chromosome arrangement inheritance is that the order of chromosomes changes completely after one cell division, as the orange line 1 indicates.

Based on the data and results presented in Chapter 3, one semi-conservative model is presented as the blue curve 3 in Figure 5.7. This curve describes the gradual changes of chromosome arrangement during cell divisions. After 5 or 6 generations, the dissimilarity will arrive at the level of the chromosome dissimilarity in non-related cells. As chromosome positions in the cell nucleus are not random, the dissimilarity of random chromosome arrangement is higher than the chromosome dissimilarity in non-related cells.

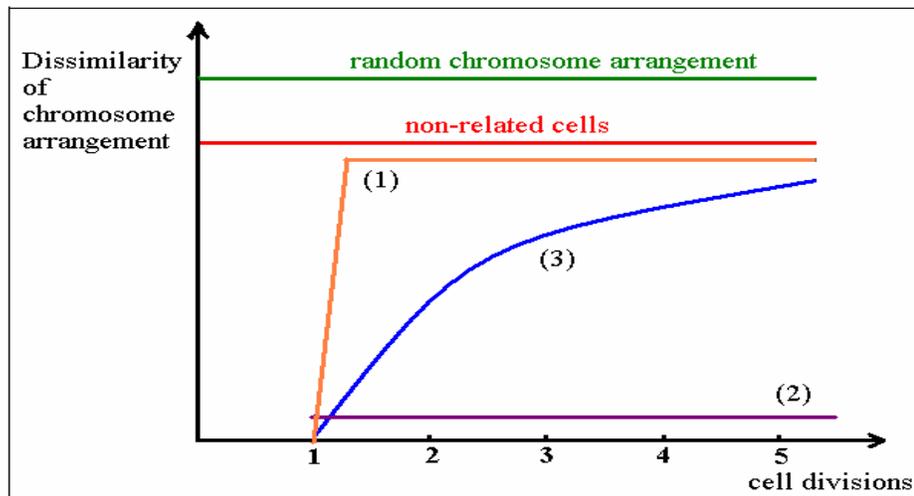


Figure 5.7 Semi-conservative model for the chromosome arrangement inheritance during cell divisions.

5.3 Future direction: to develop new quantitative methods for chromatin spatial distribution at three levels

5.3.1 The integration of spatial organization quantitative methods and genome sequence

The combination of human genome sequencing and the technical advances in microscopy and molecular cytogenetics has changed the approaches to study cancer genetics, and made computation almost an equal to the more conventional molecular tools (Mellman 2004).

Recently, the "-omic" strategies like genomic and proteomic has made us aware of the convoluted nature of biological systems. From genome-sequence data researchers can understand that genes and regulatory sequences are organized linearly on chromosomes. However, from sequence data, it is difficult to understand how the sequences are organized in the cell nucleus, and how spatial and temporal genome organization contributes to gene regulation.

New quantitative methods should be developed on three different levels: a. replication domain/gene level, b. chromosome level, and c. genome level. These new computational tools should be user-friendly, fast enough to realize automatic (or semi-automatic) analysis. If possible, these new methods should be able to be extended easily for quantification of 4D (3D + time) or 5D (4D + different channels) data. The development of reliable methods to quantify the genome spatial organization will help in the diagnosis of human diseases.

So the new challenge in the future will be to integrate these approaches and genome sequences at different levels, using the new methods in Systems Biology and Computational Cytogenetics, to yield a full picture of normal cells, leukemia cells and other cancer cells.

5.3.2 Quantification at three different levels (replication domain/genes, chromosomes, genome)

5.3.2.1 Replication domain/gene level

Genes can be found anywhere within a chromosome territory, regardless of their transcriptional activity (Mahy et al. 2002, Kurz et al. 1996). The replication domains are ~1 Mb in size and constitute a distinct level of chromosome territory organization, as they are maintained during consecutive cell cycles (Jackson et al. 1998, Zink et al. 1999). Based on the techniques presented in this thesis, the spatial distribution study of different genes and replication domains using quantitative methods will have a significant effect on our understanding of genome expression.

5.3.2.2 Chromosome level

The natural unit of subdivision of the genome is the chromosome. With tracking and other quantitative methods, the movement and dynamics of single chromosomes during mitosis or different cell divisions can be described in details.

5.3.2.3 Genome level

During mitosis, the global chromosome movements should be considered in the following four processes:

In the prophase, the genome was packaged into very compact structures without changing their positions relative to each other (Visser et al. 2000, Chubb et al. 2002). Sequentially metaphase chromosomes attached to spindle microtubules at their kinetochores in a bipolar or stochastic manner (Taniura et al. 1995). Metaphase chromosomes then move as individual units during mitosis to form the metaphase plate. In the anaphase, sister chromatids move individually towards opposite poles of the spindle to form daughter nuclei. The decondensation of chromatin fully reconstitutes interphase nuclei in daughter cells (Georgatos 2001, Andrulis et al. 2001).

How is the genome spatially and temporally organized in different species? The development of quantitative methods based on geometric spatial transformation at three different levels will be helpful to answer this question.

It is important to note that there are close relationships between these three different levels. For example, broken ends of different chromosomal regions can join only if they are in close spatial proximity. The territories of chromosomes 9 and 22 are in close vicinity, and are associated in pairs detected as false-positive ABL-BCR fusion, so the

high frequency of t(9,22) in human haemoblastoses is related to the close neighborhood of the ABL and BCR genes in the nuclei of the critical cell lines (Kozubek et al. 1999).

5.3.3 Improvement of biological experiment design and image processing techniques

The improvement of biological experiment design and image processing techniques is important to develop new computational tools and to detect the mechanism of gene expression. For instance, as a universal fluorescent marker, GFP can be fused to many proteins to visualize virtually any cellular structure in the environment of the living cell (Chalfie et al. 1994, Heun et al. 2001), although for a long time cellular structures have been investigated mostly in fixed specimens. 3D cell and tissue culture systems can be used instead of traditional cell culture systems. The latest image processing systems can already support "Live Digital Image Processing", which means that numerous real-time functions can be executed during image acquisition. These functions include online histogram calculation which is used to monitor image brightness and contrast. These new techniques will make real-time quantitative analysis of live cell nuclear organization possible.

Reference

Abney J. R., Cutler B., Fillbach M. L., Axelrod D., and Scalettar B. A. 1997. Chromatin dynamics in interphase nuclei and its implications for nuclear structure. *J. Cell Biol.* 137:1459-1468.

Abranches R., Beven A. F., Aragon-Alcaide L. and Shaw P. J. 1998. Transcription sites are not correlated with chromosome territories in wheat nuclei. *J. Cell Biol.* 143:5-12.

Akao Y., and Isobe M. 2000. Molecular analysis of the rearranged genome and chimeric mRNAs caused by the t(6;11)(q27;q23) chromosome translocation involving MLL in an infant acute monocytic leukemia. *Genes Chromosomes Cancer.* 27:412-7.

Andrulis E. D., Neiman A. M., Zappulla D. C., and Sternglanz R. 1998. Perinuclear localization of chromatin facilitates transcriptional silencing. *Nature.* 394:592-595.

Belongie S., Malik J. and Puzicha J. 2002. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.* 24:509-522.

Bickmore W. A., and Chubb J. R. 2003. Chromosome position: Now where was I? *Curr. Biol.* 13:R357-R359.

Bloomfield C. D., Goldman A., Hossfeld D., and Chapelle A. de la. 1984. Fourth International Workshop on Chromosomes in Leukemia 1982: clinical significance of chromosomal abnormalities in acute nonlymphoblastic leukemia. *Cancer Genet. Cytogenet.* 11:332-350.

Bolzer A., Kreth G., Solovei I., Koehler D., Saracoglu K., Fauth C., Müller S., Eils R., Cremer C., Speicher M. R. and Cremer T. 2005. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. In press.

Bookstein F. L. 1989. Principal wraps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Mach. Intell.* 11:567-585.

Bookstein F. L. 1991. *Morphometric Tools for Landmark Data, Geometry and Biology.* Cambridge University Press. 455 pp.

Bornfleth H., Edelmann P., Zink D., Cremer T., and Cremer C. 1999. Quantitative motion analysis of subchromosomal foci in living cells using four-dimensional microscopy. *Biophys. J.* 77:2871-2886.

Boveri T. 1909. Die blastomerenkerne von *Ascaris megalocephala* und die Theorie der Chromosomenindividualitat. *Arch. fuer Zellforsch.* 3:181-268.

Boyle S., Gilchrist S., Bridger J. M., Mahy N. L., Ellis J. A., and Bickmore W. A. 2001. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum. Mol. Genet.* 10:211-219.

Bridger J. M., Boyle S., Kill I. R., and Bickmore W. A. 2000. Re-modelling of nuclear architecture in quiescent and senescent human fibroblasts. *Curr. Biol.* 10:149-152.

Brown K. E., Amoils S., Horn J. M., Buckle V. J., Higgs D. R., Merkenschlager M., and Fisher A. G. 2001. Expression of alpha- and beta-globin genes occurs within different nuclear domains in haemopoietic cells. *Nat. Cell Biol.* 3:602-606.

Brown K. E., Baxter J., Graf D., Merkenschlager M., and Fisher A. G. 1999. Dynamic repositioning of genes in the nucleus of lymphocytes preparing for cell division. *Mol. Cell.* 3:207-217.

Brown K. E., Guest S. S., Smale S. T., Hahm K., Merkenschlager M., and Fisher A. G. 1997. Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell.* 91:845-854.

Brown L. G. 1992. A survey of image registration techniques. *ACM Comput. Surveys.* 24:325-376.

Budd T. A. 2001. *An Introduction to Object-Oriented Programming*, 3rd edition. Addison-Wesley, MA. 648 pp.

Bursen A., Moritz S., Gaussmann A., Moritz S., Dingermann T., and Marschalek R. 2004. Interaction of AF4 wild-type and AF4-MLL fusion protein with SIAH proteins: indication for t(4;11) pathobiology? *Oncogene.* 23:6237-6249.

Carmo-Fonseca M. 2002. Understanding nuclear order. *Trends Biochem. Sci.* 27:332-334.

Chalfie M., Tu Y., Euskirchen G., Ward W. W., and Prasher D. C. 1994. Green fluorescent protein as a marker for gene expression. *Science.* 263:802-805.

Chambeyron S., and Bickmore W. A. 2004. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev.* 18:1119-30.

Chen C. S., Hilden J. M., Frestedt J., Domer P. H., Moore R., Korsmeyer S. J., and Kersey J. H. 1993. The chromosome 4q21 gene (AF-4/FEL) is widely expressed in normal tissues and shows breakpoint diversity in t(4;11)(q21;q23) acute leukemia. *Blood.* 82:1080-5.

Cheng L., Ramesh K. H., Radel E., Ratech H., Wei D., and Cannizzaro L. A. 2001. Characterization of t(11;19)(q23;p13.3) by fluorescence in situ hybridization analysis in a pediatric patient with therapy-related acute myelogenous leukemia. *Cancer Genet. Cytogenet.* 129:17-22.

Chubb J. R., and Bickmore W. A. 2003. Considering nuclear compartmentalization in the light of nuclear dynamics. *Cell.* 112:403-406.

Chubb J. R., Boyle S., Perry P., and Bickmore W. A. 2002. Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr. Biol.* 12:439-445.

Cline H. E., Lorensen W. E., Ludke S., Crawford C. R., and Teeter B. C. 1988. Two algorithms for the three-dimensional reconstruction of tomograms. *Med. Phys.* 15:320-327.

Comings D. E. 1980. Arrangement of chromatin in the nucleus. *Hum. Genet.* 53:131-143.

Cornbluth M. N., Greulich-Bode K. M., Loucas B. D., Arsuaga J., Vazquez M., Sachs R. K., Bruckner M., Molls M., Hahnfeldt P., Hlatky L., and Brenner D. J. 2003. Chromosomes are predominantly located randomly with respect to each other in interphase human cells. *J. Cell Biol.* 159:237-244.

Cornforth M. N., Greulich-Bode K. M., Loucas B. D., Arsuaga J., Vazquez M., Sachs R. K., Bruckner M., Molls M., Hahnfeldt P., Hlatky L. and Brenner D. J. 2002. Chromosomes are predominantly located randomly with respect to each other in interphase human cells. *J Cell Biol.* 159:237-44.

Corral J., Lavenir I., Impey H., Warren A. J., Forster A., Larson T. A., Bell S., McKenzie A. N., King G., and Rabbitts T. H. 1996. An MLL-AF9 fusion gene made by homologous recombination causes acute leukemia in chimeric mice: a method to create fusion oncogenes. *Cell.* 85:853-861.

Cremer T., Cremer C., Baumann H., Luedtke E. K., Sperling K., Teuber V., and Zorn C. 1982. Rabl's model of the interphase chromosome arrangement tested in Chinese hamster cells by premature chromosome condensation and laser UV-microbeam experiments. *Hum. Genet.* 60:46-56.

Cremer T., Kreth G., Koester H., Fink R. H., Heintzmann R., Cremer M., Solovei I., Zink D., and Cremer C. 2000. Chromosome territories, interchromatin domain compartment, and nuclear matrix: an integrated view of the functional nuclear architecture. *Crit. Rev. Eukaryot. Gene Expr.* 10:179-212.

Cremer T. and Cremer C. 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* 2:292-301.

Cremer T., Kupper K., Dietzel S., Fakan S. 2004. Higher order chromatin architecture in the cell nucleus: on the way from structure to function. *96:555-67.*

Croft J. A., Bridger J. M., Boyle S., Perry P., Teague P., and Bikmore W. A. 1999. Differences in the localization and morphology of chromosomes in the human nucleus. *J. Cell Biol.* 145:1119-1131.

Csink A. K., and Henikoff S. 1996. Genetic modification of heterochromatic association and nuclear organization in *Drosophila*. *Nature.* 381:529-531.

Csink A. K., and Henikoff S. 1998. Large-scale chromosomal movements during interphase progression in *Drosophila*. *J. Cell Biol.* 143:13-22.

Dalla-Favera R., Bregni M., Erikson J., Patterson D., Gallo R. C., and Croce C. M. 1982. Human c-MYC onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. *Proc. Natl. Acad. Sci. USA.* 79:7824-7827.

Debernardi S., Lillington D., and Young B.D. 2004. Understanding cancer at the chromosome level: 40 years of progress. *Eur. J. Cancer.* 40:1960-1967.

Dernburg A. F., Broman K. W., Fung J. C., Marshall W. F., Philips J., Agard D. A., and Sedat J. W. 1996. Perturbation of nuclear architecture by long-distance chromosome interactions. *Cell.* 85:745-759.

Dimartino J. F., and Cleary M. L. 1999. MLL rearrangements in haematological malignancies: lessons from clinical and biological studies. *Br. J. Haematol.* 106:614-26.

Djabali M., Selleri Parry P., Bower M., Young B. D., and Evans G. A. 1992. A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute leukaemias. *Nat. Genet.* 2:113-118.

Dobson C. L., Warren A. J., Pannell R., Forster A., Lavenir I., Corral J., Smith A. J., and Rabbitts T. 1999.

The MLL-AF9 gene fusion in mice controls myeloproliferation and specifies acute myeloid leukaemogenesis. *EMBO J.* 18:3564-3574.

Domer P. H., Fakharzadeh S. S., Chen C. S., Jockel J., Johansen L., Silverman G. A., Kersey J. H., and Korsmeyer S. J. 1993. Acute mixed-lineage leukemia t(4;11)(q21;q23) generates an MLL-AF4 fusion product. *Proc. Natl. Acad. Sci. USA.* 90:7884-8.

Driel R. van, and Fransz P. 2004. Nuclear architecture and genome functioning in plants and animals: what can we learn from both? *Exp. Cell Res.* 296:86-90.

Dryden I. L. and Mardia K. V. 1998. *Statistical shape analysis.* John Wiley and Sons Ltd. 376 pp.

Duchon J. 1976. Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces. *RAIRO Analyse Numerique.* 10:5-12.

Dundr M., and Misteli T. 2001. Functional architecture in the cell nucleus. *Biochem. J.* 356:297-310.

Eckel B. 2002. *Thinking in Java, 3rd edition.* Prentice Hall PTR. 1119 pp.

Eils R., Gerlich D., Tvarusko W., Spector D. L., and Misteli T. 2000. Quantitative imaging of pre-mRNA splicing factors in living cells. *Mol. Biol. Cell.* 11:413-418.

Eils R., Uhrig S., Saracoglu K., Satzler K., Bolzer A., Petersen I., Chassery J., Ganser M., and Speicher M. R. 1998. An optimized, fully automated system for fast and accurate identification of chromosomal rearrangements by multiplex-FISH (M-FISH). *Cytogenet. Cell Genet.* 82:160-71.

Eils R., Dietzel S., Bertin E., Schrock E., Speicher M. R., Ried T., Robert-Nicoud M., Cremer C. and Cremer T. 1996. Three-dimensional reconstruction of painted human interphase chromosomes: active and inactive X chromosome territories have similar volumes but differ in shape and surface structure. *J. Cell Biol.* 135:1427-40.

Elliott B., and Jasin M. 2002. Double-strand breaks and translocations in cancer. *Cell. Mol. Life Sci.* 59:373-385.

Ferrant M., Nabavi A., Kikinis R., and Warfield S. K. 2001. Real-time simulation and visualization of volumetric brain deformation for image guided neurosurgery. *SPIE Med. Imaging.* 4319:366-373.

Feuerbach F., Galy V., Trelles-Sticken E., Fromont-Racine M., Jacquier A., Gilson E., Olivo-Marin J. C.,

Scherthan H., and Nehrbass U. 2002. Nuclear architecture and spatial positioning help establish transcriptional states of telomeres in yeast. *Nat. Cell Biol.* 4:214-221.

Fieres J. 2001. A new Point Set Registration Algorithm (Diploma thesis, Heidelberg University, Germany).

Fieres J., Mattes J., and Eils R. 2001. Automatic Deformable Surface Registration using Thin-Plate Splines: Matching local features while preserving global contiguity. In DAGM (Die Deutsche Arbeitsgemeinschaft für Mustererkennung). Springer-Verlag, Berlin. 2191:76-83.

Fraleigh J. B. 1996. *Calculus with Analytic Geometry* (section 18.5.4), 9th edition. Addison-Wesley Publishing Company. 1139 pp.

Francastel C., Schubeler D., Martin D. I., and Groudine M. 2000. Nuclear compartmentalization and gene activity. *Nat. Rev. Mol. Cell Biol.* 1:137-143.

Francastel C., Walters M. C., Groudine M., and Martin D. I. 1999. A functional enhancer suppresses silencing of a transgene and prevents its localization close to centromeric heterochromatin. *Cell.* 99:259-269.

Franke U. 1981. High-resolution ideograms of trypsin-Giemsa banded human chromosomes. *Cytogenet Cell Genet.* 31:24-32.

Frestedt J. L., Hilden J. M., Moore R. O., and Kersey J. H. 1996. Differential expression of AF4/FEL mRNA in human tissues. *Genet. Anal.* 12:147-149.

Fukuhara S., Rowley J. D., Variakojis D., and Golomb H. M. 1979. Chromosome abnormalities in poorly differentiated lymphocytic leukemia. *Cancer Res.* 39:3119-3128.

Galy V., Olivo-Marin J. C., Scherthan H., Doye V., Rascalou N., and Nehrbass U. 2000. Nuclear pore complexes in the organization of silent telomeric chromatin. *Nature.* 403:108-112.

Gamboa-Aldeco A., Fellingham L. L., and Chen G. T. Y. 1986. Correlation of 3D surfaces from multiple modalities in medical imaging. In *Proceedings Medicine XIV/PACS IV*, SPIE. 626:467-473.

Gao J., Köhler D., Eils R., Solovei I., Cremer T., and Mattes J. 2004. Assessing The Similarity of Spatial Configurations Using Distance Differences and Bending Energy: Application To Chromosomal Interphase

Arrangements In HeLa Cell Clones. In IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'2004), Arlington, VA, USA. 1400-1403.

Gasser S. M. 2002. Visualizing chromatin dynamics in interphase nuclei. *Science*. 296:1412-1416.

Gasser S. M., and Laemmli U. K. 1986. Cohabitation of scaffold binding regions with upstream/enhancer elements of three developmentally regulated genes of *D. melanogaster*. *Cell*. 46:521-530.

Gebhard M., Eils R., and Mattes J. 2002. Segmentation of 3D objects using NURBS surfaces for quantification of surface and volume dynamics. In International Conference on Diagnostic Imaging and Analysis (ICDIA'2002). Shanghai, China. 125-130.

Gebhard M., Mattes J., and Eils R. 2001. An active contour model for segmentation based on cubic B-splines and Gradient Vector Flow. In Medical Image Computing and Computer-Assisted Intervention (MICCAI'2001). Springer-Verlag. 2208:1373-1375.

Georgatos S.D. 2001. The inner nuclear membrane: simple, or very complex?. *Embo J*. 20:2989–2994.

Gerlich D., Beaudouin J., Gebhard M., Ellenberg J., and Eils R. 2001. Four-dimensional imaging and quantitative reconstruction to analyse complex spatiotemporal processes in live cells. *Nat. Cell. Biol.* 3:852-855.

Gerlich D., Beaudouin J., Kalbfuss B., Daigle N., Eils R., and Ellenberg J. 2003 (A). Global chromosome positions are transmitted though mitosis in mammalian cells. *Cell*. 112:751-764.

Gerlich D., and Ellenberg J. 2003 (B). 4D imaging to assay complex dynamics in live specimens. *Nat. Cell. Biol.* 5:S14-S19.

Gerlich D., and Ellenberg J. 2003 (C). Dynamics of chromosome positioning during the cell cycle. *Curr. Opin. Cell Biol.* 15:664-671.

Gerlich D., Mattes J., and Eils R. 2003 (D). Quantitative motion analysis and visualization of cellular structures. *Methods*. 29:3-13.

Germain F., Doisy A., Ronot X., and Tracqui P. 1999. Characterization of cell deformation and migration using a parametric estimation of image motion. *IEEE Trans. Biomed. Eng.* 46:584-600.

Gilbert N., Boyle S., Fiegler H., Woodfine K., Carter N. P., and Bickmore W. A. 2004 (B). Chromatin Architecture of the Human Genome: Gene-Rich Domains Are Enriched in Open Chromatin Fibers. *Cell*. 118:555-566.

Gilbert N., Gilchrist S., and Bickmore W. A. 2004 (B). Chromatin organization in the Mammalian nucleus. *Int. Rev. Cytol.* 242:283-336.

Goldberg M., Harel A., Brandeis M., Rechsteiner T., Richmond T. J., Weiss A. M., and Gruenbaum Y. 1999. The tail domain of lamin Dm0 binds histones H2A and H2B. *Proc. Natl. Acad. Sci. USA.* 96:2852-2857.

Goldman R. D., Gruenbaum Y., Moir R. D., Shumaker D. K., and Spann T. P. 2002. Nuclear lamins: building blocks of nuclear architecture. *Genes Dev.* 16:533-547.

Gonzalez R. C., and Woods R. E. 1992. Digital image processing. Addison Wesley Publishing Company. 716 pp.

Gu Y., Cimino G., Alder H., Nakamura T., Prasad R., Canaani O., Moir D. T., Jones C., Nowell P. C., Croce C. M., and Canaani E. 1992. The (4;11)(q21;q23) chromosome translocations in acute leukemias involve the VDJ recombinase. *Proc. Natl Acad. Sci. USA.* 89:10464-10468.

Gustashaw, K.M. 1991. ChromosomeStains. The ACT Cytogenetics Laboratory Manual, Second Edition, edited by M.J. Barch. The Association of Cytogenetic Technologists, Raven Press, Ltd., New York.

Haaf T., and Schmid M. 1991. Chromosome topology in mammalian interphase nuclei. *Exp. Cell Res.* 192:325-32.

Habermann F. A., Cremer M., Walter J., Kreth G., von Hase J., Bauer K., Wienberg J., Cremer C., Cremer T., and Solovei I. 2001. Arrangements of macro- and microchromosomes in chicken cells. *Chromosome Res.* 9:569-584.

Harder R., and Desmarais R. 1972. Interpolation using surface splines. *J. Aircraft.* 9:189-191.

Harris J. W., and Stocker H. 1998. Handbook of Mathematics and Computational Science. Springer-Verlag, New York, NY. 1056 pp.

Heitz E. 1932. Die herkunft der chromocentren. *Planta.* 18:571-635.

Heun P., Laroche T., Shimada K., Furrer P., and Gasser S. M. 2001. Chromosome dynamics in the yeast interphase nucleus. *Trends Cell Biol.* 11:519-525.

Hochstrasser M., Mathog D., Gruenbaum Y., Saumweber H., and Sedat J. W. 1986. Spatial organization of chromosomes in the salivary gland nuclei of *Drosophila melanogaster*. *J. Cell Biol.* 102:112-123.

Hollander M., and Wolfe D. A. 1973, *Nonparametric Statistical Inference*. John Wiley and Sons, Inc., New York, NY.

Hudson M. M., Raimondi S. C., Behm F. G., and Pui C. H. 1991. Childhood acute leukemia with t(11;19)(q23;q3). *Leukemia.* 5:1064-1068.

Iida S., Seto M., Yamamoto K., Komatsu H., Tojo A., Asano S., Kamada N., Ariyoshi Y., Takahashi T., and Ueda R. 1993. MLLT3 gene on 9p22 involved in t(9;11) leukemia encodes a serine/proline rich protein homologous to MLLT1 on 19p13. *Oncogene.* 8:3085-92.

Mellman I. and Misteli T. 2004. Computational cell biology. *J. Cell Biol.* 161:463-464.

Ishii K., Arib G., Lin C., Van Houwe G., and Laemmli U. K. 2002. Chromatin boundaries in budding yeast: the nuclear pore connection. *Cell.* 109:551-562.

Jaffe E. S., Harris N. L., Stein H., and Vardiman J. W. (eds). 2001. *World Health Organization Classification of Tumours, Pathology and Genetics: Tumours of the Hematopoietic and Lymphoid Tissues*. In International Agency for Research on Cancer (IARC), Lyon, France.

Jahne B., and Jahne B. 1997. *Digital Image Processing: Concepts, Algorithms, and Scientific Applications*. Springer-Verlag, Berlin. 555 pp.

Jin Q. W., Fuchs J., and Loidl J. 2000. Centromere clustering is a major determinant of yeast interphase nuclear organization. *J. Cell Sci.* 113:1903-1912.

Joh T., Hosokawa Y., Suzuki R., Takahashi T., and Seto M. 1999. Establishment of an inducible expression system of chimeric MLL-LTG9 protein and inhibition of Hox a7, Hox b7 and Hox c9 expression by MLL-LTG9 in 32Dcl3 cells. *Oncogene.* 18:1125-1130.

Joh T., Yamamoto K., Kagami Y., Kakuda H., Sato T., Yamamoto T., Takahashi T., Ueda R., Kaibuchi K.,

and Seto M. 1997. Chimeric MLL products with a Ras binding cytoplasmic protein AF6 involved in t(6;11) (q27;q23) leukemia localize in the nucleus. *Oncogene*. 15:1681-7.

Kay A. C. 1993. The Early History of Smalltalk. In *The Second ACM SIGPLAN History of Programming Languages Conference (HOPL-II)*. ACM SIGPLAN Notices. 28:69-75.

Knudson A. G. 2001. Two genetic hits (more or less) to cancer. *Nat. Rev. Cancer*. 1:157-162.

Köhler D. 2002. Variabilität oder Konstanz der Chromosomenanordnung während des klonalen Wachstums menschlicher Zellen (Diplomarbeit, University of Munich, Germany).

Köhler D., Gao J., Mattes J., Eils R., Cremer T., and Solovei I. 2004 (A). Arrangements of Chromosome Territories in Cloned Human Cells. Annual Convention of the German Society for Cell Biology, Berlin, Germany. For abstract: *Eur. J. of Cell Biol.* Vol. 83, Suppl. 54, pp 60, ISSN 0171-9335.

Köhler D., Gao J., Mattes J., Eils R., Cremer T., and Solovei I. 2004 (B). Changes in the arrangement of chromosome territories in human cell clones. The 15th International Chromosome Conference (ICC XV '2004), London, United Kingdom.

Kolmogorov A. N. 1933. Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*. 4: 83-91.

Koss L. G. 1998. Characteristics of chromosomes in polarized normal human bronchial cells provide a blueprint for nuclear organization. *Cytogenet. Cell Genet.* 82: 230-237.

Kozubek S., Lukasova E., Mareckova A., Skalnikova M., Kozubek M., Bartova E., Kroha V., Krahulcova E., and Slotova J. 1999. The topological organization of chromosomes 9 and 22 in cell nuclei has a determinative role in the induction of t(9,22) translocation and in the pathogenesis of t(9,22) leukemias. *Chromosoma*. 108:426-435.

Kreth G., Finsterle J., von Hase J., Cremer M., and Cremer C. 2004. 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.* 86:2803-12.

Landau L. D., and Lifshitz E. M. 1959. *Theory of Elasticity*. Butterworth-Heinemann. 195 pp.

Langer T., Metzler M., Reinhardt D., Viehmann S., Borkhardt A., Reichel M., Stanulla M., Schrappe M.,

Creutzig U., Ritter J., Leis T., Jacobs U., Harbott J., Beck J. D., Rascher W., and Repp R. 2003. Analysis of t(9;11) Chromosomal Breakpoint Sequences in Childhood Acute Leukemia: Almost Identical MLL Breakpoints in Therapy-Related AML After Treatment Without Etoposides. *Genes Chromosomes Cancer*. 36:393-401.

Laroche T., Martin S. G., Gotta M., Gorham H. C., Pryde F. E., Louis E. J., and Gasser S. M. 1998. Mutation of yeast Ku genes disrupts the subnuclear organization of telomeres. *Curr. Biol*. 8:653-656.

Lavau C., Du C., Thirman M., and Zeleznik-Le N. 2000. Chromatin-related properties of CBP fused to MLL generate a myelodysplastic-like syndrome that evolves into myeloid leukemia. *EMBO J*. 19:4655-4664.

Lavau C., Szilvassy S. J., Slany R., and Cleary M. L. 1997. immortalization and leukemic transformation of a myelomonocytic precursor by retrovirally transduced HRX-ENL. *EMBO J*. 16:4226-4237.

Lemon B., and Tjian R. 2000. Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev*. 14:2551-2569.

Lichter P., Cremer T., Borden J., Manuelidis L., and Ward D. C. 1988. Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Hum. Genet*. 80:224-234.

Luderus E., den Blaauwen J. L., de Smit O. J., Compton D. A., and van Driel R. 1994. MBinding of matrix attachment regions to lamin polymers involves single-stranded regions and the minor groove. *Mol. Cell Biol*. 14:6297-6305.

Lukasova E., Kozubek S., Kozubek M., Kjeronska J., Ryznar L., Horakova J., Krahulcova E., and Horneck G. 1997. Localisation and distance between ABL and BCR genes in interphase nuclei of bone marrow cells of control donors and patients with chronic myeloid leukaemia. *Hum. Genet*. 100:525-535.

Luo R. T., Lavau C., Du C., Simone F., Polak P. E., Kawamata S., and Thirman M. J. 2001. The Elongation Domain of ELL Is Dispensable but Its ELL-Associated Factor 1 Interaction Domain Is Essential for MLL-ELL-Induced Leukemogenesis. *Mol. Cell Biol*. 21:5678-5687.

Mahy N. L., Perry P. E., and Bickmore W. A. 2002. Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *J. Cell Biol*. 159:753-763.

- Manders E. M., Kimura H., and Cook P. R. 1999. Direct imaging of DNA in living cells reveals the dynamics of chromosome formation. *J. Cell Biol.* 144:813-821.
- Manuelidis L. 1985. Indications of centromere movement during interphase and differentiation. *Ann. N. Y. Acad. Sci.* 450:205-221.
- Manuelidis L. 1985. Individual interphase chromosome domains revealed by in-situ hybridization. *Hum. Genet.* 71:288-293.
- Marshall W. F. 2002. Order and disorder in the nucleus. *Curr. Biol.* 12:R185-R192.
- Marshall W. F., Straight A., Marko J. F., Swedlow J., Dernburg A., Belmont A. S., Murray A. W., Agard D. A., and Sedat J. W. 1997. Interphase chromosomes undergo constrained diffusional motion in living cells. *Curr. Biol.* 7:930-939.
- Mattes J., Fieres J., and Eils R. 2002. A shape adapted motion model for non-rigid registration. In *SPIE Medical Imaging 2002: Image Processing*. San Diego, CA. 4684:518-527.
- Meyer B. 2000. Object-oriented software construction. 2nd edition. Prentice Hall PTR. 1296 pp.
- Misteli T. 2001. Protein dynamics: implications for nuclear architecture. *Science.* 291:843-847.
- Misteli T. 2004. Spatial positioning: a new dimension in genome function. *Cell.* 119:153-156.
- Misteli T., Caceres J. F., and Spector D. L. 1997. The dynamics of a pre-mRNA splicing factor in living cells. *Nature.* 387:523-527.
- Mitelman F. 1981. The third international workshop on chromosomes in leukemia. *Cancer Genet. Cytogenet.* 4:96-98.
- Mitelman F. 2000. Recurrent chromosome aberrations in cancer. *Mutat. Res.* 462:247-253.
- Mitterbauer G., Zimmer C., Pire-Danoewinata H., Haas O. A., Hojas S., Schwarzinger I., Greinix H., Jager U., Lechner K., and Mannhalter C. 2000. Monitoring of minimal residual disease in patients with MLL-AF6-positive acute myeloid leukaemia by reverse transcriptase polymerase chain reaction. *Br. J. Haematol.* 109:622-8.

Miyoshi H., Shimizu K., Kozu T., Maseki N., Kaneko Y., and Ohki M. 1991. t(8;21) breakpoints on chromosome 21 in acute myeloid leukemia are clustered within a limited region of a single gene, AML1. *Proc. Natl. Acad. Sci. USA.* 88:10431-4.

Moir R. D., Yoon M., Khuon S., and Goldman R. D. 2000. Nuclear lamins A and B1: different pathways of assembly during nuclear envelope formation in living cells. *J. Cell Biol.* 151:1155 -1168.

Muratani M., Gerlich D., Janicki S. M., Gebhard M., Eils R., and Spector D. L. 2002. Metabolic-energy-dependent movement of PML bodies within the mammalian cell nucleus. *Nat. Cell. Biol.* 4:106-110.

Murmann A. 2004. Topology of genes in mammalian cell nuclei with special emphasis on the MLL gene and its translocation partners (PhD thesis, Heidelberg University, Germany).

Nagele R. G., Freeman T., McMorro L., Lee H. Y. 1995. Precise spatial positioning of chromosomes during prometaphase: evidence for chromosomal order. *Science* 270: 1831-1835.

Nagele R. G., Freeman T., Fazekas J., Lee K. M., Thomson Z., et al. 1998. Chromosome spatial order in human cells: evidence for early origin and faithful propagation. *Chromosoma* 107: 330-338.

Neves H., Ramos C., da Silva M. G., Parreira A., and Parreira L. 1999. The nuclear topography of ABL, BCR, PML, and RARalpha genes: evidence for gene proximity in specific phases of the cell cycle and stages of hematopoietic differentiation. *Blood.* 93:1197-1207.

Nikiforova M. N., Stringer J. R., Blough R., Medvedovic M., Fagin J. A., and Nikiforov Y. E. 2000. Proximity of chromosomal loci that participate in radiation-induced rearrangements in human cells. *Science.* 290:138-141.

Nilson I., Lochner K., Siegler G., Greil J., Beck J. D., Fey G. H., and Marschalek R. 1996. Exon/intron structure of the human ALL-1 (MLL) gene involved in translocations to chromosomal region 11q23 and acute leukaemias. *Br. J. Haematol.* 93:966-972.

Nilson I., Reichel M., Ennas M. G., Greim R., Knorr C., Siegler G., Greil J., Fey G. H., and Marschalek R. 1997. Exon/intron structure of the human AF-4 gene, a member of the AF-4/LAF-4/FMR-2 gene family coding for a nuclear protein with structural alterations in acute leukaemia. *Br. J. Haematol.* 98:157-169.

O'Brien T. P., Bult C. J., Cremer C., Grunze M., Knowles B. B., Langowski G., McNally J., Pederson T.,

Politz J. C., Pombo A., Schmahl G., Spatz J. P., and van Driel R. 2003. Genome function and nuclear architecture: from gene expression to nanoscience. *Genome Res.* 13:1029-1041.

Parada L. A., McQueen P. G., Munson P. J., and Misteli T. 2002 (A). Conservation of relative chromosome positioning in normal and cancer cells. *Curr. Biol.* 12:1692-7.

Parada L. A., and Misteli T. 2002 (B). Chromosome positioning in the interphase nucleus. *Trends Cell Biol.* 12:425-432.

Parada L. A., Roix J. J., and Misteli T. 2003. An uncertainty principle in chromosome positioning. *Trends Cell Biol.* 13:393-396.

Pederson T. 2001. Protein mobility in the nucleus: what are the right moves? *Cell.* 104:635-638.

Pederson T. 2002. Dynamics and genome-centricity of interchromatin domains in the nucleus. *Nat. Cell Biol.* 4:E287-E291.

Pederson T. 2003. Gene territories and cancer. *Nat. Genet.* 34:242-243.

Pederson T. 2004. The spatial organization of the genome in mammalian cells. *Curr. Opin. Genet. Dev.* 14:203-209.

Pelizzari C. A., Chen G. T., Spelbring D. R., Weichselbaum R. R., and Chen C. T. 1989. Accurate three-dimensional registration of CT, PET and/or MR images of the brain. *J. Comput. Assist. Tomogr.* 13:20-26.

Phair R. D., and Misteli T. 2001. Kinetic modelling approaches to in vivo imaging. *Nat. Rev. Mol. Cell Biol.* 2:898-907.

Pinkel D., Straume T., and Gray J. W. 1986. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc. Natl. Acad. Sci. USA.* 83:2934-2938.

Platani M., Goldberg I., Lamond A. I., and Swedlow J. R. 2002. Cajal body dynamics and association with chromatin are ATP-dependent. *Nat. Cell Biol.* 4:502-508.

Platani M., Goldberg I., Swedlow J. R., and Lamond A. I. 2000. In vivo analysis of Cajal body movement, separation, and joining in live human cells. *J. Cell Biol.* 151:1561-1574.

Pombo A., Jones E., Iborra F. J., Kimura H., Sugaya K., Cook P. R., and Jackson D. A. 2000. Specialized transcription factories within mammalian nuclei. *Crit. Rev. Eukaryot. Gene Expr.* 10:21-9.

Prasad R., Gu Y., Alder H., Nakamura T., Canaani O., Saito H., Huebner K., Gale R. P., Nowell P. C., Kuriyama K., and et al. 1993. Cloning of the ALL-1 fusion partner, the AF-6 gene, involved in acute myeloid leukemias with the t(6;11) chromosome translocation. *Cancer Res.* 53:5624-8.

Prasad R., Yano T., Sorio C., Nakamura T., Rallapalli R., Gu Y., Leshkowitz D., Croce C. M., and Canaani E. 1995. Domains with transcriptional regulatory activity within the ALL1 and AF4 proteins involved in acute leukemia. *Proc. Natl. Acad. Sci. USA.* 92:12160-12164.

Press W. H., Teukolsky S. A., Vetterling W. T., and Flannery B. P. 1990. *Numerical recipes in C: The Art of Scientific Computing.* Cambridge University Press. 1020 pp.

R Development Core Team. 2004. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rabl C. 1885. Über Zellteilung. In: C. Gegenbauer, Editor, *Morphologisches Jahrbuch.* 10: 214-258.

Rangarajan A., Chui H., Mjolsness E., Pappu S., Davachi L., Goldman-Rakic P. S., and Duncan J. 1997. A robust point matching algorithm for autoradiograph alignment. *Med. Image Anal.* 4:379-398.

Roix J. J., McQuenn P. G., Munson P. J., Parada L. A., and Misteli T. 2003. Spatial proximity of translocation-prone gene loci in human lymphomas. *Nat. Genet.* 34:287-291.

Rowley J. D. 2001. Chromosome translocations: dangerous liaisons revisited. *Nat. Rev. Cancer.* 1:245-250.

Rowley J.D. 1998. The critical role of chromosome translocations in human leukemias. *Annu. Rev. Genet.* 32:495-519.

Rubnitz J. E., Morrissey J., Savage P. A., and Cleary M. L. 1994. ENL, the gene fused with HRX in t(11;19) leukemias, encodes a nuclear protein with transcriptional activation potential in lymphoid and myeloid cells. *Blood.* 84:1747-1752.

Sachs R. K., Hahnfeld P., and Brenner D. J. 1997. Proximity effects in the production of chromosome aberrations by ionizing radiation. *Int. J. Radiat. Biol.* 71:1-19.

Schardin M., Cremer T., Hager H. D., and Lang M. 1985. Specific staining of human chromosomes in Chinese hamster × man hybrid cell lines demonstrates interphase chromosome territories. *Hum. Genet.* 71:281-287.

Schermelleh L., Solovei I., Zink D., Cremer T. 2001. Two-color fluorescence labeling of early and mid-to-late replicating chromatin in living cells. *Chromosome Res.* 9:77-80.

Shannon M. F. 2003. A nuclear address with influence. *Nat. Genet.* 34:4-6.

Shilatifard A., Duan D. R., Haque D., Florence C., Schubach W. H., Conaway J. W., and Conaway R. C. 1997. ELL2, a new member of an ELL family of RNA polymerase II elongation factors. *Proc. Natl. Acad. Sci. USA.* 94:3639-3643.

Shilatifard A., Lane W. S., Jackson K. W., Conaway R. C., and Conaway J. W. 1996. An RNA polymerase II elongation factor encoded by the human ELL gene. *Science.* 271:1873-1876.

Shtivelman E., Lifschitz B., Gale R. P., and Canaani E. 1985. Fused transcript of ABL and BCR genes in chronic myelogenous leukaemia. *Nature.* 315:550-554.

Spector D. L. 2001. Nuclear domains. *J. Cell Sci.* 114:2891-2893.

Spector D. L. 2003. The dynamics of chromosome organization and gene regulation. *Annu. Rev. Biochem.* 72:573-608.

Stoffler D., Fahrenkrog B., and Aebi U. 1999. The nuclear pore complex: from molecular architecture to functional dynamics. *Curr. Opin. Cell Biol.* 11:391-401.

Strahl B. D., and Allis C. D. 2000. The language of covalent histone modifications. *Nature.* 403:41-45.

Strissel P. L., Strick R., Tonek R. J., Roe B. A., Rowley J. D., and Zeleznik-Le N. J. 2000. DNA structural properties of AF9 are similar to MLL and could act as recombination hot spots resulting in MLL/AF9 translocations and leukemogenesis. *Human Mol. Genet.* 9:1671-1679.

Sturtevant A. H. 1925. The effects of unequal crossing over at the bar locus in *Drosophila*. *Genetics.* 10:117-147.

Sumner A. T. 2003. *Chromosomes: Organization and Function*. Blackwell Publishing company. 384 pp.

Sun H. B., and Yokota H. 1999. Correlated positioning of homologous chromosomes in daughter fibroblast cells. *Chromosome Res.* 7:603-610.

Super H. G., Strissel P. L., Sobulo O. M., Burian D., Reshmi S. C., Roe B., Zeleznik-Le N. J., Diaz M. O., and Rowley J. D. 1997. Identification of Complex Genomic Breakpoint Junctions in the t(9;11) MLL-AF9 Fusion Gene in Acute Leukemia. *Genes Chromosomes Cancer.* 20:185-195.

Suzuki S., Chiba K., Toyoshima N., Kurosawa M., Hashino S., Musashi M., and Asaka M. 2001. Chronic eosinophilic leukemia with t(6;11)(q27;q23) translocation. *Ann. Hematol.* 80:553-6.

Tanabe H., S.M., M. Neusser, J. von Hase, E. Calcagno, M. Cremer, I. Solovei, C. Cremer, and T. Cremer. 2002. Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc. Natl. Acad. Sci. USA.* 99:4424-4429.

Tanabe S., Zeleznik-Le N. J., Kobayashi H., Vignon C., Espinosa R. 3rd., LeBeau M. M., Thirman M. J., and Rowley J. D. 1996. Analysis of the t(6;11)(q27;q23) in leukemia shows a consistent breakpoint in AF6 in three patients and in the ML-2 cell line. *Genes Chromosomes Cancer.* 15:206-16.

Taniura H., Glass C., and Gerace L. 1995. A chromatin binding site in the tail domain of nuclear lamins that interacts with core histones. *J. Cell Biol.* 131:33-44.

Taylor R. H., Lavallée S., Burdea G. C., and Mösges R. 1995. *Computer Integrated Surgery: Technology and Clinical Applications.* MIT Press, Cambridge, MA. 736 pp.

Thirman M. J., Diskin E. B., Bin S. S., Ip H. S., Miller J. M., and Simon M. C. 1997. Developmental analysis and subcellular localization of the murine homologue of ELL. *Proc. Natl. Acad. Sci. USA.* 94:1408-1413.

Thirman M. J., Gill H. J., Burnett R. C., Mbangkollo D., McCabe N. R., Kobayashi H., Ziemins-Van Der Poel S., Kaneko Y., Morgan R., Sandberg A. A., Chaganti R. S. K., Larson R. A., Le Beau M. M., Diaz M. O., and Rowley J. D. 1993. Rearrangement of the MLL gene in acute lymphoblastic and acute myeloid leukemias with 11q23 chromosomal translocations. *N. Engl. J. Med.* 329:909-914.

Thirman M. J., Levitan D. A., Kobayashi H., Simon M. C., and Rowley J. D. 1994. Cloning of ELL, a gene that fuses to MLL in a t(11;19)(q23;p13.1) in acute myeloid leukemia. *Proc. Natl. Acad. Sci. USA.* 91:12110-12114.

Thompson D. W. 1917. *On Growth and Form.* Dover Publications. 1116 pp.

Thomson I., Gilchrist S., Bickmore W. A., and Chubb J. R. 2004. The Radial Positioning of Chromatin Is Not Inherited through Mitosis but Is Established De Novo in Early G1. *Curr. Biol.* 14:166-172.

Tkachuk D. C., Kohler S., and Cleary M. L. 1992. Involvement of a homolog of *Drosophila trithorax* by 11q23 chromosomal translocation in acute leukemias. *Cell.* 71:691-700.

Trick R., Strissel P. L., Borgers S., Smith S. L., and Rowley J. D. 2000. Dietary bioflavonoids induce cleavage in the MLL gene and may contribute to infant leukemia. *Proc. Natl. Acad. Sci. USA*:4790-4795.

Tsujimoto Y., Finger L. R., Yunis J., Nowell P. C., and Croce C. M. 1984. Cloning of the chromosome breakpoint of neoplastic B cells with the t(14;18) chromosome translocation. *Science.* 266.

Tvaruskó W., Bentele M., Misteli T., Rudolf R., Kaether C., Spector D. L., Gerdes H. H., and Eils R. 1999. Time-resolved analysis and visualization of dynamic processes in living cells. *Proc. Natl. Acad. Sci. USA.* 96:7950-7955.

Uckun F. M., Herman-Hatten K., Crotty M. L., Sensel M. G., Sather H. N., Tuel-Ahlgren L., Sarquis M. B., Bostrom B., Nachman J. B., Steinherz P. G., Gaynon P. S., and Heerema N. 1998. Clinical significance of MLL-AF4 fusion transcript expression in the absence of a cytogenetically detectable t(4;11)(q21;q23) chromosomal translocation. *Blood.* 92:810-821.

Vazquez J., Belmont A. S., and Sedat J. W. 2001. Multiple regimes of constrained chromosome motion are regulated in the interphase *Drosophila* nucleus. *Curr. Biol.* 11:1227-1239.

Vig B. K. 1981. Sequence of centromere separation: analysis of mitotic chromosomes in man. *Hum. Genet.* 57:247-252.

Visser A. E., and Aten J. A. 1999. Chromosomes as well as chromosomal subdomains constitute distinct units in interphase nuclei. *J. Cell Sci.* 112:3353-3360.

Visser A. E., Eils R., Jauch A., Little G., Bakker P. J., Cremer T., and Aten J. A. 1998. Spatial distributions of early and late replicating chromatin in interphase chromosome territories. *Exp. Cell Res.* 243:398-407.

Visser A. E., Jaunin F., Fakan S., and Aten J. A. 2000. High resolution analysis of interphase chromosome domains. *J. Cell Sci.* 113:2585-2593.

Vliet L. J. van, and Verbeek P. W. 1993. Curvature and bending energy in digitized 2D and 3D images. In 8th Scandinavian Conference on Image Analysis (SCIA '93), Tromso, Norway. 1403-1410.

Volpi E. V., Chevret E., Jones T., Vatcheva R., Williamson J., Beck S., Campbell R. D., Goldsworthy M., Powis S. H., Ragoussis J., Trowsdale J., and Sheer D. 2000. Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J. Cell Sci.* 113:1565-1576.

Walter J., Schermelleh L., Cremer M., Tashiro S., and Cremer T. 2003. Chromosome order in HeLa cells changes during mitosis and early G1, but is stably maintained during subsequent interphase stages. *J. Cell Biol.* 160:685-697.

Wiedeman L. M., MacGregor A., and Caldas C. 1999. Analysis of the region of the 5' end of the MLL gene involved in genomic duplication events. *Br. J. Haematol.* 105:256-264.

Wilkie G. S., Shermoen A. W., O'Farrell P. H. and Davis I. 1999. Transcribed genes are localized according to chromosomal position within polarized *Drosophila* embryonic nuclei. *Curr. Biol.* 9:1263-1266.

Williams R. R., and Fisher A. G. 2003. Chromosomes, positions please! *Nat. Cell Biol.* 5:388-390.

Williams R.R., Broad S., Sheer D., and Ragoussis J. 2002. Subchromosomal positioning of the epidermal differentiation complex (EDC) in keratinocyte and lymphoblast interphase nuclei. *Exp. Cell Res.* 272:163-175.

Yong I. T., Walker J. E., and Bowie J. E. 1974. An analysis technique for biological shape I. *Info Control.* 25:357-370.

Zech L., Haglund U., Nilsson K., and Klein G. 1976. Characteristic chromosomal abnormalities in biopsies and lymphoid-cell lines from patients with Burkitt and non-Burkitt lymphomas. *Int. J. Cancer.* 17:47-56.

Zeisig B. B., Garcia-Cuellar M. P., Winkler T. H., and Slany R. K. 2003. The oncoprotein MLL-ENL disturbs hematopoietic lineage determination and transforms a biphenotypic lymphoid/myeloid cell. *Oncogene.* 22:1629-37.

Ziemin-Van Der Poel S., McCabe N. R., Gill H. J., Espinosa III R., Patel Y., Harden A., Rubinelli P., Smith S. D., LeBeau M. M., Rowley J. D., and Diaz M. O. 1991. Identification of a gene, MLL, that spans the breakpoint in 11q23 translocations associated with human leukemias. *Proc. Natl. Acad. Sci. USA.* 88:10735-10739.

Zickler D. and Kleckner N. 1999. Meiotic chromosomes: integrating structure and function. *Annu. Rev. Genet.* 33:603-754.

Zink D., Cremer T., Saffrich R., Fischer R., Trendelenburg M. F., Ansorge W., and Stelzer E. H. 1998. Structure and dynamics of human interphase chromosome territories in vivo. *Hum. Genet.* 102:241-251.

Zink D., Amaral M. D., Englmann A., Lang S., Clarke L. A., Rudolph C., Alt F., Luther K., Braz C., Sadoni N., Rosenecker J., and Schindelhauer D. 2004 (A). Transcription-dependent spatial arrangements of CFTR and adjacent genes in human cell nuclei. *J. Cell Biol.* 166:815-825.

Zink D., Fischer A. H. and Nickerson J. A. 2004 (B). Nuclear structure in cancer cells. *Nat. Rev. Cancer.* 4:677-687.

Abbreviations

2D	two dimensional
3D	three dimensional
<i>AF4</i>	ALL1 fused gene from chromosome 4
<i>AF6</i>	ALL1 fused gene from chromosome 6
<i>AF9</i>	ALL1 fused gene from chromosome 9
AL	acute leukemia
ALL	acute lymphoblastic leukemia
AML	acute Myeloid (myeloblastic) leukemia
bcr	breakpoint cluster region
bp	base pair
CML	chronic myelogenous leukemia
cDNA	complementary deoxyribonucleic acid
chr	normal chromosome
DAPI	4', 6-diamidin-2'-phenylindol-dihydrochloride
del	chromosome with deletion
der	derivative chromosome
DNA	deoxyribonucleic acid
<i>E. coli</i>	<i>Escherichia coli</i>
<i>ELL</i>	eleven nineteen lysine-rich leukemia gene on chromosome 19
<i>ENL</i>	eleven nineteen leukemia gene on chromosome 19
FISH	fluorescence in situ hybridization
G-banding	Giemsa-banding
GFP	green fluorescent protein
HSA	chromosomes from <i>Homo sapiens</i> (human)
IV-file	file in Inventor format
kb	kilo base
kDa	kilo dalton

m	meter
M-phase	mitotic phase of the cell cycle
<i>M. Muntjak</i>	Muntiacus muntjak
<i>M. reevesi</i>	Muntiacus reevesi
MMV	Muntiacus muntjak chromosome
MRE	Muntiacus reevesi chromosome
rDNA	ribosomal DNA
t-AML	treatment-related AML
T-cell	T lymphocyte

Publication List

Papers

1. Juntao Gao, Daniela Köhler, Irina Solvei, Thomas Cremer, Roland Eils, Julian Mattes. Assessing The Similarity of Spatial Configurations Using Distance Differences and Bending Energy: Application To Chromosomal Interphase Arrangements In HeLa Cell Clones, IEEE ISBI 2004, pp1400-1403.(Proceedings of the 2004 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Arlington, VA, USA, April 15-18, 2004. IEEE ISBI 2004)
2. Juntao Gao, A. E. Murmann, J. D. Rowley, P. Lichter, R. Eils. Three-dimensional quantitative tools to analyze the spatial arrangement of translocation partner genes. In preparation.
3. A. E. Murmann, Juntao Gao, M. Encinosa, M. E. Peter, R. Eils, P. Lichter, J. D. Rowley. Gene density within 2 Mbp of a locus determines its 3D position in the interphase nucleus of hematopoietic cells. In preparation.
4. Daniela Köhler, Juntao Gao, Roland Eils, Thomas Cremer, Julian Mattes, Irina Solovej. Inheritance and changes of chromosome arrangement in proliferating human cells. In preparation.

Poster/Oral Presentations

1. Daniela Köhler, Juntao Gao, Julian Mattes, Roland Eils, Thomas Cremer, Irina Solovej: Arrangements of Chromosome Territories in Cloned Human Cells.

Poster presentation, Annual Convention of the German Society for Cell Biology, Berlin, Germany, March 24-27, 2004. For abstract: European Journal of Cell Biology, Vol. 83, Suppl. 54, pp 60, March 2004, ISSN 0171-9335.

2. Andrea Murmann, Juntao Gao: New computational tools to analyze the spatial distribution of nuclear entities.

Oral Presentation. [FOM 2004](#) (Focus On Microscopy, Philadelphia, USA, April 4-7, 2004).

3. Daniela Köhler, Juntao Gao, Julian Mattes, Roland Eils, Thomas Cremer, Irina Solovei: Changes in the arrangement of chromosome territories in human cell clones.

Oral presentation, [Gordon Research Conferences](#), Molecular Cytogenetics, July 18-23 2004, The Queen's College, Oxford University, UK.

4. Juntao Gao, Andrea Murmann: New 3D computational tools to evaluate and visualise the spatial arrangement of potential translocation partner genes in fixed nuclei.

Oral presentation (presented by C. Bacher), [ICC XV](#) (The 15th International Chromosome Conference, London, United Kingdom, 5th - 10th September, 2004).

5. Daniela Köhler, Juntao Gao, Julian Mattes, Roland Eils, Thomas Cremer, Irina Solovei: Changes in the arrangement of chromosome territories in human cell clones.

Poster presentation, [ICC XV](#) (The 15th International Chromosome Conference, London, United Kingdom, 5th - 10th September, 2004).

6. Daniela Köhler, Juntao Gao, Julian Mattes, Roland Eils, Thomas Cremer, Irina Solovei: Arrangements of Chromosome Territories in Human Cell Clones.

Oral presentation, Marie Curie Conferences and Training Courses on arrayCGH and Molecular Cytogenetics, First Conference. Wellcome Trust Sanger Institute, Cambridge, 29.Sep.- 02.Oct. 2004.

Acknowledgement

I would like to express my gratitude to the many individuals whose invaluable contributions have made this thesis possible.

Firstly, I thank our group leader, Prof. Dr. Roland Eils, who made this PhD work possible and offered me enough scientific freedom to accomplish my work here with diverse collaborators.

I am indebted to Dr. Julian Mattes for his work in obtaining the financial support for my PhD research work and for his countless suggestions on the calculation details in Chapter 2 and 3 during our twenty months of close collaboration. Despite the great distance after his relocation in Austria, we managed to carry on a cooperation to complete the project in Chapter 3 and the discussion about rewriting of whole Chapter 2 of my PhD thesis.

I thank Prof. Dr. Thomas Cremer of LMU in Munich, who initiated the project in Chapter 3, for the discussions of the quantitative calculations in Chapter 3. I owe my thanks to Daniela Köhler, a PhD student of LMU in Munich for the production of the needed data which were pivotal for the project in Chapter 3 and the discussion involved in this project. During the last two years, we cooperated closely on the details of the project. In addition, I wish to extend my appreciation and thanks to Dr. Irina Solovei for her coordination of this project.

I am deeply grateful to Andrea Murmann, whose valuable data, opinions and discussions for the project of translocation genes played an important role in the completion of this project in Chapter 4. Also, I benefited from Prof. Dr. Peter Lichter of DKFZ in Germany and Prof. Dr. Janet D. Rowley of University of Chicago in USA for initiating the project in Chapter 4 and providing me an opportunity to work on this challenging topic.

I would like to extend my sincere thanks to our BMCV subgroup leader, Prof. Dr. Karl Rohr, who revised the technical part of this PhD thesis.

I owe my thanks to Dr. Rittgen in the department of Biostatistics in DKFZ and Dr. Mansmann in the department of Medicine Biostatistics in UniversitaetKlinikum Heidelberg for the statistical discussion and calculation. I would like to thank Johannes Fieres, Kirchhoff-Institut für Physik of the Universitaet Heidelberg for his contributions to calculate the bending energy.

In addition, I wish to extend my warmest thanks to all the group members in our lab (especially Karl-Heinz Gross, Peter Weyrich and Siwei Yang), for their support and all the scientific truth that they instilled in me.