The Korean Communications in Statistics
Vol. 13 No. 1, 2006, pp. 167-175

# Applications of Geostatistics to the Quantitative Analysis of Genetic Instability in Carcinogenesis[1]

Hyoung-Moon Kim[2]

## Abstract

It has long been recognized that cancer is a genetic disease. To find this measures of genetic instability, stain cells with chromosome specific probes using chromosome in-situ hybridization technique is adopted. Even though in-situ hybridization technique is powerful, truncation of nuclei often results in under-representation of chromosome copies in slides due to the sectioning of tissue blocks. Because of this problem we suggest three different methods to analyze the cervical cancer data set. We observe that genetic instability is an increasing function of histology and our suggested model is the best in detecting genetic instability of tumorigenesis processes.

*Keywords* : Tumorigenesis; Geostatistics; Smoothing; Random-coefficient model.

## 1. Introduction

Cell division (proliferation) is a physiological process that occurs in almost all tissues and under many circumstances. Normally homeostasis, the balance between proliferation and programmed cell death, usually in the form of apoptosis, is maintained by tightly regulating both processes to ensure the integrity of organs and tissues. Mutations in DNA that lead to cancer disrupt these orderly processes by disrupting the program regulating the processes.

Carcinogenesis (Andreeff and Pinkel, 1999) is caused by this mutation of the genetic material of normal cells, which upsets the normal balance between proliferation and cell death. This results in uncontrolled cell division and tumor formation. This uncontrolled cell division is called genetic instability. The uncontrolled and often rapid proliferation of cells can lead to benign tumors; some types of these may turn into malignant tumors (cancer). Benign tumors do not spread to other parts of the body or invade other tissues, and they are rarely a threat to life unless they compress vital structures or are physiologically active

1) I thank to Jack Lee(Biostatistics Department, MDAnderson) for providing the cervical cancer data set.

2) Assistant professor, Department of Applied Statistics, Konkuk University, Seoul 143-701, Korea. E-mail : hmk966a@yahoo.co.kr

(for instance, producing a hormone). Malignant tumors can invade other organs, spread to distant locations (metastasize) and become life threatening.

More than one mutation is necessary for carcinogenesis. In fact, a series of several mutations to certain classes of genes is usually required before a normal cell will transform into a cancer cell. Only mutations in those certain types of genes which play vital roles in cell division, cell death, and DNA repair will cause a cell to lose control of its proliferation.

In this article, to find this measures of genetic instability, stain cells with chromosome specific probes using chromosome in-situ hybridization technique is adopted (Kim et al, 1993; Hittelman, 2001). This technique involves the use of DNA probes that recognize either chromosome-specific repetitive target sequences or sequences along the whole chromosome length or chromosome segments. Although in-situ hybridization technique is powerful, truncation of nuclei often results in under-representation of chromosome copies in slides due to the sectioning of tissue blocks. Because of this problem we suggest three different methods to analyze the cervical cancer data set.

The rest of an article is organized as follows. Section 2 provides data description. Section 3 proposes three different methods and their results. We observe that genetic instability is an increasing function of histology and our suggested model is the best in detecting genetic instability of tumorigenesis processes. A discussion is drawn in Section 4.

## 2. Data Description

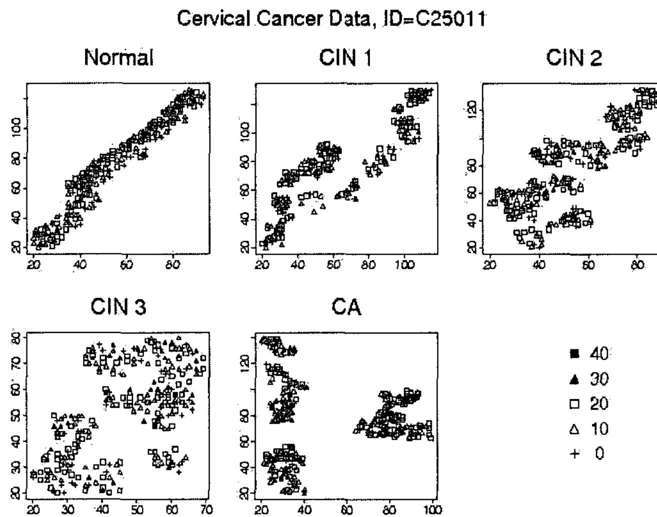
There are 28 eligible patients in this study. A total of 77 tissue samples of different stages of carcinogenesis(NORM, CIN1, CIN2, CIN3, CA) are obtained from these patients. Histological progression is proceeded from Normal epithelium to CIN 1, CIN 2, CIN 3, and Cancer. <Table 1> is the summary of samples by histology. <Table 2> is the summary of diagnosis of these 28 patients.

<Table 1> 77 Samples by Histology

| Histology | Number of Samples |
|---|---|
| NORM | 21 |
| CIN1 | 15 |
| CIN2 | 15 |
| CIN3 | 20 |
| CA | 6 |
| Total | 77 |

<Table 2> Summary of the Patients Diagnosis

| Diagnosis | Number of Patients |
|---|---|
| CA | 11 |
| CIS | 12 |
| CIN3 | 5 |
| Total | 28 |




<Figure 1> Cervical Cancer Data for a Patient ID=C25011

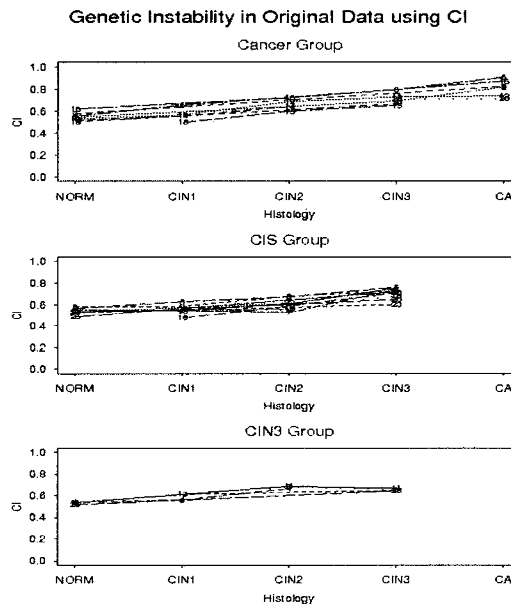
One patient has all stages so we plot it in <Figure 1> as an example of the data sets. The values (0-40) in <Figure 1> denote the observed chromosome copies $z_i \times 10$ at location $(x_i, y_i), i = 1, \ldots, n_i$. Since the coordinates of each data set have different ranges, we standardized the range of them to [0,1]. So there is no relative spatial information change. Basically our hypothesis of interest is that multistep tumorigenesis process is an increasing function of genetic instability and degree of genetic instability could be identified between the clinical diagnosis. To check hypotheses we will employ three different approaches at the following section.

## 3. Detection of Genetic Instability

This section proposes three different approaches to detect multistep

tumorigenesis progress using chromosome index, chromosome polysomy, and smoothed chromosome polysomy.

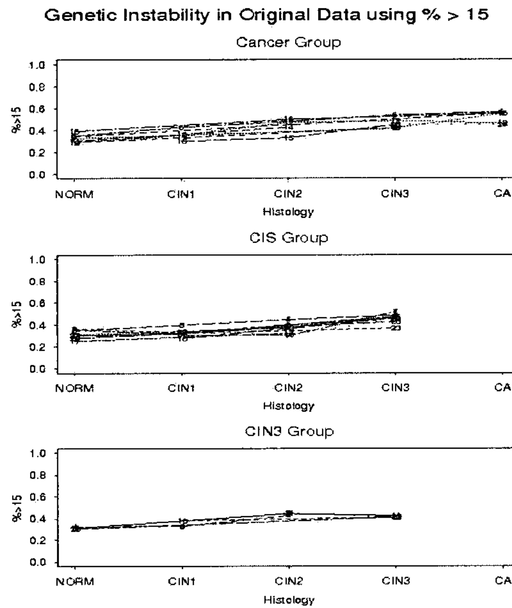
The first method uses chromosome index (Brambilla and Brambilla, 1999) which is the average number of chromosome copy per experimental cell divided by the average number of chromosome copy in normal cell. We plot this in <Figure 2>. Each patient is connected with different line types and this also applied to <Figure 3> and <Figure 5>. Since portions of nuclei were cut in the preparation of tissue sections, the average number of labeled spots per nucleus in normal cells was less than the expected 20. This point is a reason why we develop the following 2 procedures.




<Figure 2> Genetic Instability in Original Data using CI(Chromosome Index)

The second method applies chromosome polysomy which is a proportion of data points greater than some predefined cutoff point. We use 15 as the cutoff point which is plotted in <Figure 3>.

Final method uses smoothed chromosome polysomy. The weak point of the first and the second method is that the values are integers so they are insensitive to detect instability of tumorigenesis process. The other weak point of the above two methods is that they ignore spatial information residing in tissue samples(Cressie, 1993). To overcome these weaknesses, we first smooth each tissue sample using nearest neighbor method(Besag, 1974).




<Figure 3> Genetic Instability in Original Data Using %>15

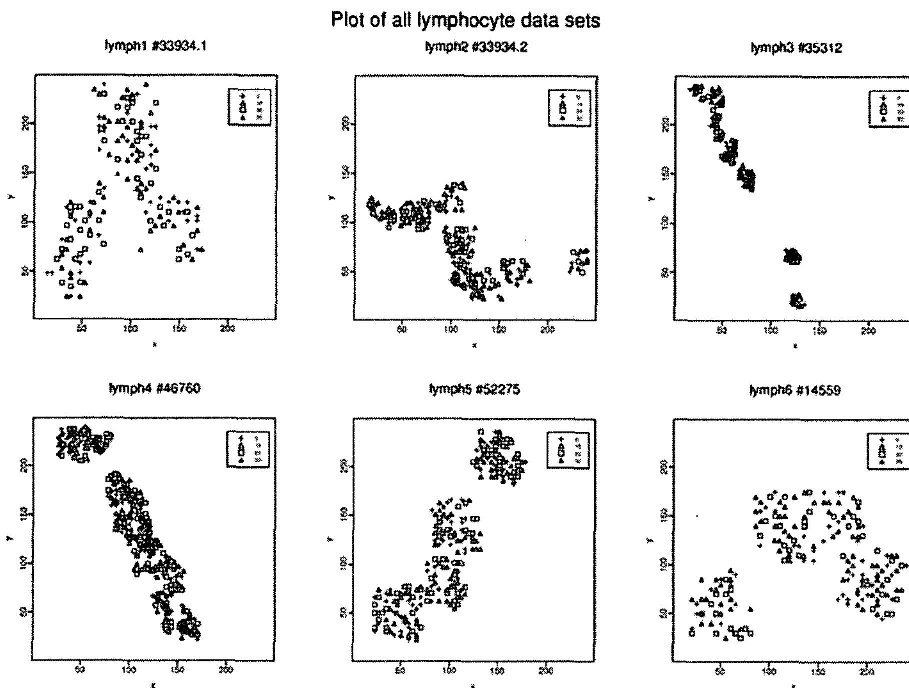
We use first and second order nearest neighborhood schemes. First order scheme uses 4 sample points nearest to each point, then arithmetic mean of 5 points including point itself is a smoothed chromosome polysomy. We might use a kernel method or weighted mean instead of arithmetic average. Second order scheme uses 8 sample points nearest to each point. Furthermore we also incorporated maximum distance of all cells. That is all cells within each neighborhood scheme should be within a maximum distance $d$. To find this we first fit each data set using spherical variogram(Cressie, 1993),

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ c_0 + c\left(3h/(2a) - h^3/(2a^3)\right), & 0 < h < a \\ c_0 + c, & h \geq a \end{cases}$$

where $a$ is range, $c_0$ is called nugget, and $c_0 + c$ is sill value. $h$ is called separation or lag. This variogram model is most widely used in spatial analysis. The maximum distance is then set to median of all ranges, 0.273, for a robust purpose, since the range contains the information on how far data are correlated in space. We might use mark variogram instead of (geostatistical) variogram that used in this article. However it will be enough to use a (geostatistical) variogram when we estimate a range since it is a property of long-range variability(Stoyan and Walder, 2000; Diggle, 2003). So we will stick to use (geostatistical) variogram.

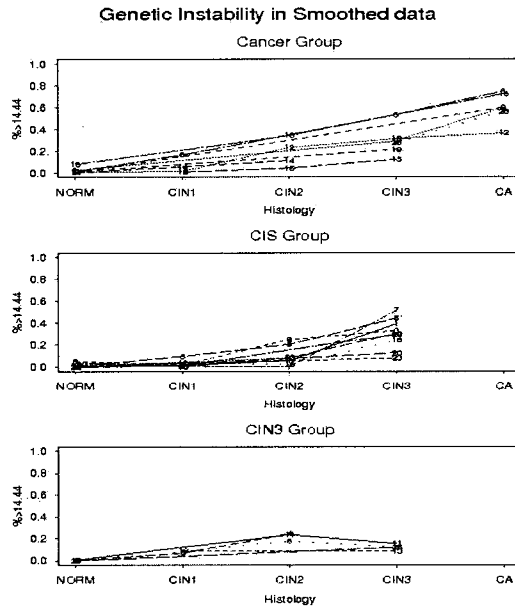
After smoothing, we need to find an appropriate cutoff point to detect instability

of cells. To achieve this six lymphocyte data sets for the normal cells were used and plotted in <Figure 4>. Legends are the same as those of <Figure 1>. We follow the same steps as genetic instability data set. 95 percentiles from the distribution of nearest neighbor means of all lymphocyte data sets are obtained. They will be used as the cutoff points since they contain the information of what is abnormal for the normal data set. 95 percentiles of the first and second order nearest neighborhood schemes are 15 and 14.44. The second order nearest neighborhood with this cutoff 14.44 are plotted in <Figure 5>. The other plot is not presented because it is similar to <Figure 5>.

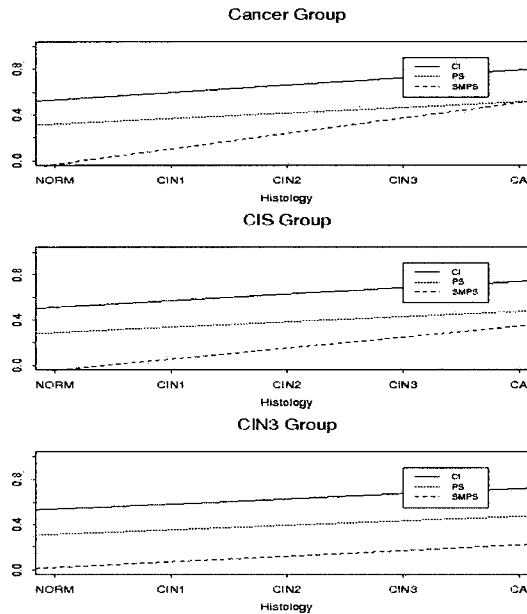



<Figure 4> Six lymphocyte data sets

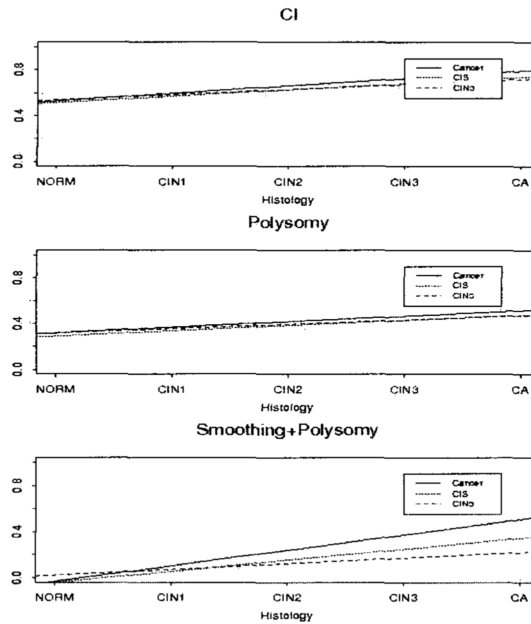
Looking at three plots(<Figures 2, 3, 5>) we observe that genetic instability is an increasing function of histology for all three plots. However the slopes of graphs are a little different. So we fit the random-coefficient regression model for the above three plots (Littell et al., 1996). We test the hypothesis that slopes are the same among 3 groups(Cancer, CIS, and CIN3 groups). P-values for this hypothesis are 0.1281, 0.5914, and 0.0028 for each method which reveals that the final method is the best in detecting genetic instability of tumorigenesis processes. We plot the fitted model in <Figure 6> by each group and in <Figure 7> by three different methods.




<Figure 5> Genetic Instability in Smoothed Data




<Figure 6> Fitted models by group: CI denotes Chromosome Index, PS denotes Polysomy, and SMPS denotes Smoothing+Polysomy




<Figure 7> Fitted models by method: CI denotes Chromosome Index, PS denotes Polysomy, and SMPS denotes Smoothing+Polysomy

## 4. Discussion

We found that chromosome in-situ hybridization is useful for detecting genomic instability on tissue sections. The nuclei truncation problem caused by tissue section can be addressed by nearest neighbour averaging. Genetic instability increases with progressive histology in the multi-step tumorigenesis process.

## 참고문헌

[1] Andreeff, M. and Pinkel, D. (1999). *Introduction to Fluorescence In Situ Hybridization: Principles and Clinical Applications*, Wiley-Liss, Inc.

[2] Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society Ser. B*, Vol. 36, 192-225.

[3] Brambilla, C. and Brambilla, E. (1999). *Lung Tumors*, Marcel Dekker, New York.

[4] Cressie, N.A.C. (1993). *Statistics for Spatial Data(Revised Edition)*, John Wiley and Sons. New York.

[5] Diggle, P.J. (2003). *Statistical Analysis of Spatial Point Patterns(second edition)*, Oxford University Press Inc., New York.

[6] Kim, S.Y., Lee, J.S., Ro, J.Y., Gay, M.L., Hong, W.K., and Hittelman, W.N. (1993). Interphase Cytogenetics in Paraffin Sections of Lung Tumors by Non-Isotopic in-Situ Hybridization. *American Journal of Pathology*, Vol. 142, 307-317.

[7] Hittelman, W.N. (2001). Genetic Instability in Epithelial Tissues at Risk for Cancer. *Annals of the New York Academy of Sciences*, Vol. 952, 1-12.

[8] Littell, R.C., Milliken, G.A., Stroup, W.W., and Wo, R.D. (1996). *SAS System for Mixed Models*, SAS Publishing.

[9] Stoyan, D. and Walder, O. (2000). On Variogram in Point Process Statistics, II:Models of Markings and Ecological Interpretation. *Biometrical Journal*, Vol. 42, 171-187.

[ Received March 2006, Accepted March 2006 ]