Automated Processing of 2-D Gel Electrophoretograms of Genomic DNA for Hunting Pathogenic DNA Molecular Changes

Katsutoshi Takahashi [1]
sltaka@jaist.ac.jp

Masayuki Nakazawa [2]
nakazawa@kanazawa-tc.ac.jp

Yasuo Watanabe [3]
watanabe@infor.kanazawa-it.ac.jp

Akihiko Konagaya [1]
kona@jaist.ac.jp

[1] Japan Advanced Institute of Science and Technology (JAIST) Hokuriku, Asahidai 1-1, Tatsunokuchi, Ishikawa 923-1211, Japan
[2] Kanazawa Technical College, Hisayasu 2-270, Kanazawa, Ishikawa 921-8601, Japan
[3] Kanazawa Institute of Technology, Ohgigaoka 7-1, Nonoichi, Ishikawa 921-8501, Japan

Abstract

We have developed the automated processing algorithms for 2-dimensional (2-D) electrophoretograms of genomic DNA based on RLGS (Restriction Landmark Genomic Scanning) method, which scans the restriction enzyme recognition sites as the landmark and maps them onto a 2-D electrophoresis gel. Our powerful processing algorithms realize the automated spot recognition from RLGS electrophoretograms and the automated comparison of a huge number of such images. In the final stage of the automated processing, a master spot pattern, on which all the spots in the RLGS images are mapped at once, can be obtained. The spot pattern variations which seemed to be specific to the pathogenic DNA molecular changes can be easily detected by simply looking over the master spot pattern. When we applied our algorithms to the analysis of 33 RLGS images derived from human colon tissues, we successfully detected several colon tumor specific spot pattern changes.

1 Introduction

The complete sequence of human genomic DNA is predicted to be solved by early in the next century. In the next stage of the human genome project, main research topics should shift from DNA sequence itself to (i) the identification of all of the 10^5 genes and their functions, and (ii) the regulation mechanism of genetic network.

Sometimes, the unknown function of a new gene can be identified from the genotype-phenotype correlation revealed by some sort of gene disruption and overexpression experiments. However, in practice, it is prohibited to apply such experiments to human individuals, and therefore it is very difficult to approach the function of human genes from such point of view. One of the most promising way to approach human genetic function is to analyze genetic and DNA sequential information concerning with diseases, such as cancer or genetic diseases, which seemed to be caused by specific DNA molecular changes. Such DNA molecular changes can be identified by comparing huge number of gathered genetic and DNA sequential information and then, by detecting common difference between diseased and non-diseased DNA molecules.

Restriction Landmark Genomic Scanning (RLGS) method [5, 6] is one of the experimental techniques by which one can detect individual DNA molecular changes, such as deletions, additions or amplifications. By applying RLGS 2-D gel electrophoresis to a genomic DNA, restriction enzyme recognition sites locating on the DNA can be expanded onto a 2-D electrophoresis gel as shown in Fig. 1. In this figure, *NotI*, which is a DNA methylation sensitive restriction enzyme, recognition sites on the genomic DNA extracted from colon tumor tissue are labeled with radio isotope (RI)

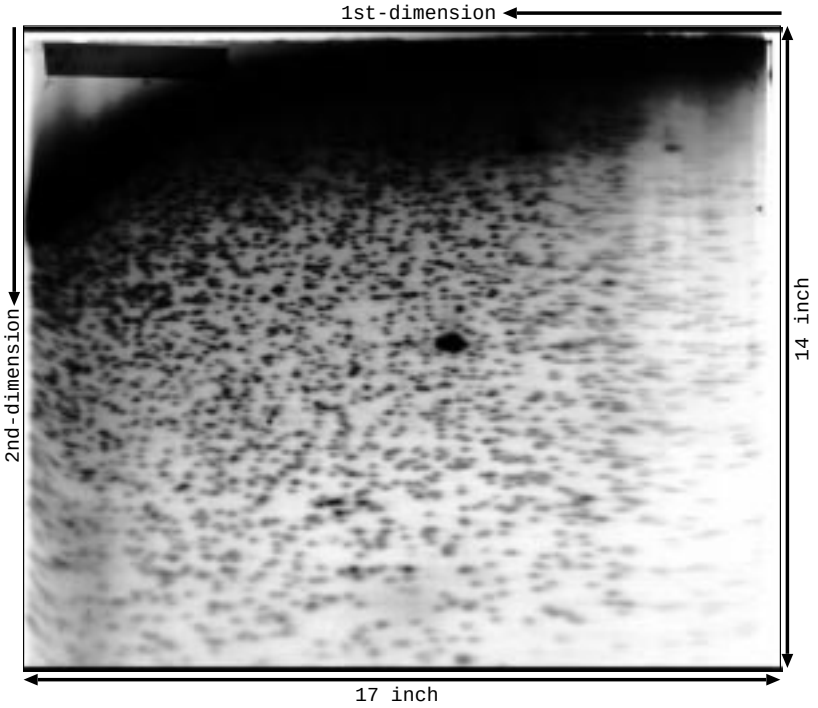


Figure 1: The RLGS 2-D gel electrophoretogram of genomic DNA extracted from colon tumor tissue.

and expanded as the RLGS 2-D gel electrophoresis image (RLGS profile), according to the following experimental procedure.

At the beginning of the experiment, the whole genomic DNA is extracted from colon tumor tissue. The purified DNA is cleaved with *NotI* (8 base recognition) and then, 3'-end-labeled with radio isotope (RI). The labeled DNA fragments are further digested by another enzyme *EcoRV* (6 base recognition) and subjected to the first dimensional agarose gel electrophoresis. After the electrophoresis, the DNA fragments are in gel digested by *HinfI* (5 base recognition), then separated by second dimensional polyacrylamide gel electrophoresis. Finally, several thousands of the landmark spots appeared on a X-ray film after auto-radiography of the dried gel.

In the RLGS profile, intensity of each spot is proportional to the copy number of corresponding restriction landmark, and the spot location reflects the relative occurrence of the recognition sites of the used enzymes. When the RLGS spot patterns (of *NotI* landmarks) are different commonly between several related RLGS profiles, it is quite possible that the DNA methylation status or copy number is changed at or near to the important regulatory region on the DNA, because the *NotI* recognition sites are frequently found at or near to the gene expression regulatory regions on genomic DNA, such as famous CpG island [2]. As same as the DNA methylation or copy number change, landmark spots translocate or disappear, or the new landmark spot appears when the DNA sequence changes at or near to a restriction landmark site.

Additionally in RLGS method, any restriction enzyme can be used for the landmark enzyme instead of *NotI*. Therefore with employing multiple different restriction enzymes, one can monitor various DNA molecular changes, which cannot be detected by other techniques such as DNA-chip and DNA-microarray [4, 9, 14, 16].

2 Previous work

As shown above, in principle, the disease specific DNA molecular changes occurring at or near to the restriction landmark sites can be detected effectively by applying RLGS method to several diseased

and non-diseased genomic DNAs, followed by precise RLGS image processing. This information gives some clue to the succeeding functional analyses and interaction analyses of the genes [1, 3, 6, 7].

We developed and reported the automated computer algorithms for RLGS electrophoresis image processing [10, 11, 15]. In this section, let us explain briefly how our algorithms process disordered RLGS profiles to extract information concerning to the DNA molecular changes occurring. In our automated algorithms, we consider a RLGS profile as a featured point pattern represented by each RLGS spot location, shape and intensity. Then, we distinguish the equivalent pair of spots from both of two RLGS profiles by means of the point pattern matching technique.

In general, RLGS profile frequently shows several irregular features in spot shape, as can be seen in Fig. 1. Some landmark spots have long-tailed shape, and the other spots show flat shape caused by signal saturation, in addition to the strongly drifted un-uniform background patterns. Our previously reported spot recognition algorithm [11] was designed to identify such spots optimally with only a few ill-recognized spots. Additionally, it effectively recognized the 'hidden spots' which showed no peak at their location but could be found at the shoulder of neighboring large spots.

Besides the irregular spot shape, RLGS profiles take non-linear distortions, i.e., the whole patterns of the RLGS profiles do not coincide even if the profiles are derived from exactly the same genomic DNAs. Because of such irregular features, RLGS profiles cannot be compared by a simple image overlay technique. With applying our previously reported pairwise spot pattern matching algorithm [11] to such disordered two RLGS spot patterns, the equivalent pairs of spots in both profiles are distinguished accurately and promptly.

It is noteworthy that our spot recognition and pairwise spot pattern matching algorithms process RLGS profiles in fully automated manner. During the processing with our algorithms, there is no need to do boring, annoying and time-consuming interactions with computer. In the following subsections, the fully automated RLGS image processing algorithms are outlined.

2.1 Fully-automated spot recognition

The location of the landmark spots in a RLGS profile are detected by 'ring operator' instead of conventional simple peak detector and Gaussian filter, in order to avoid ill-recognition of spots which have long tail or flat shape. Once the spot location is detected, the spot shape is modeled by Gaussian-type function. This Gaussian modeling of shot shape realizes both of the accurate intensity estimation and the asymmetric tailing removal.

As the next step, the hidden spots are identified. They cannot be detected by any sort of peak detector because they show no peak at their location because of their neighboring large spots. In our algorithm, the existence of hidden or un-recognized spots is detected on the differential image calculated from the background normalized original RLGS profile and the virtial profile synthesized from the above Gaussian-modeled spot shapes. Here, their shape is also modeled by Gaussian-type functions.

In the final stage of our spot recognition algorithm, all of the Gaussian-modeling parameters are refined so as to minimize the difference between the background normalized original RLGS profile and the synthesized virtial profile. During this step, almost of all ill-recognized spots are removed automatically.

2.2 Fully-automated pairwise spot pattern comparison

As described above, RLGS profiles take asymmetric distortion which is different in position by position on each gel. In our pairwise spot pattern matching algorithm, we solved such disordered spot pattern comparison problem as the matching problem of two structured graphs.

Here, we represent a RLGS spot pattern as a Delaunay net (DN) [13] in which each node takes spot intensity and shape information as attributes. Accordingly, the equivalent spot pair search in two RLGS profiles is treated as the equivalent graph-node pair search in two DNs. We solved such

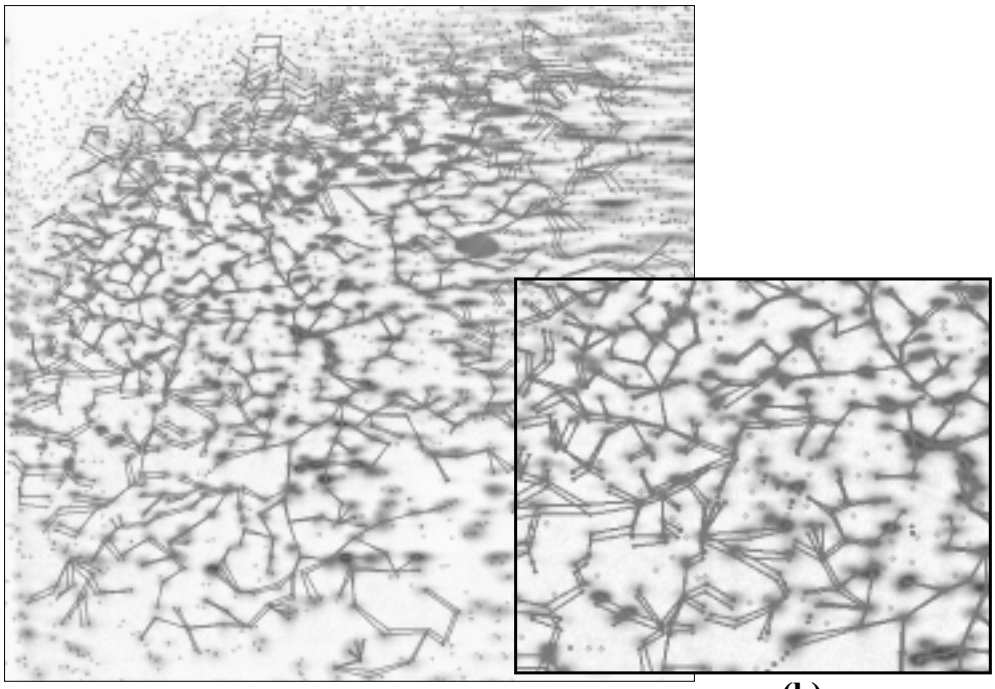


Figure 2: The equivalent RLGS landmark spots are shown overlaid on the background normalized RLGS profile of colon normal cell. (a): the landmark spots and the RNG (relative neighborhood graph) [10, 13] constructed from the matched spots are superimposed. The superimposed spots of tumor RLGS profile was depicted as the diamonds, while the spots of normal RLGS profile was drawn with square. (b): (a) was partly magnified for convenience.

graph matching problem by means of simple breadth-first graph search technique. With an initial equivalent node pair in both reference and object DNs given, the connected nodes in the reference DN are searched in breadth-first manner. During the reference DN search, equivalent node on the object DN is identified. The most suitable initial equivalent node pair in both DNs is also determined by the heuristical search technique.

With the initial equivalent node pair search and graph matching algorithms, landmark spots equivalent in two RLGS profiles can be identified accurately and rapidly, without any human interaction.

Fig. 2 demonstrates how well our automated spot recognition and pairwise spot pattern matching algorithms work. In this figure, 1066 equivalent spots were automatically identified from two RLGS profiles , which derived from a colon tumor tissue and a colon normal tissue; 2742 spots and 1879 spots were automatically recognized and Gaussian-modeled by our spot recognition algorithm, respectively.

3 Matching of multiple spot patterns

As described above, a RLGS profile takes great deals of information about the landmark restriction enzyme recognition sites locating on a genomic DNA. Of course you should compare at least two RLGS profiles and notice the difference between them to identify DNA molecular changes, such as DNA methylation change, mutation, addition, deletion or amplification. However in principle, you should compare more then two RLGS profiles and should extract only the common difference, in order to distinguish the DNA molecular changes specifically occurring on genomic DNA.

For instance, when you distinguish the DNA molecular changes specifically occurring in tumor cells, you should gather RLGS profiles derived from tumor tissue and control (normal) tissue of many patients, then compare them to identify RLGS pattern difference between tumor and control profiles. However, it is known that RLGS patterns do not coincide when the genomic DNAs are derived from different individuals or different organs, i.e., the RLGS pattern differences observed are caused by several factors. In order to separate the specific RLGS pattern difference of the tumor cell from the ones caused by individual or organic difference, you should extract only the difference common to all sets of tumor-control profiles.

It is very difficult, however, and sometimes impractical by hand to separate such specific RLGS pattern difference from the huge number sets of tumor-control profiles, even if the RLGS pattern difference of a set of tumor-control profiles is automatically detected by means of our previously reported algorithms. To overcome the difficulties in the analysis of more than two RLGS profiles, we introduced the idea of the 'master spot pattern' (MSP) of the RLGS profiles, on which all spot information of all constituent RLGS profiles are mapped, and have developed the computer algorithm for automated MSP generation. We describe the idea and the automated generation algorithm of MSP in this section.

The idea of the 'master spot pattern' is very simple. It is a sort of imaginary spot pattern which unifies the all spot information of multiple RLGS profiles at once. A master spot in a MSP is a unity of all equivalent spots found on the constituent RLGS profiles. It holds the spot intensity (and possibly the spot shape) information of the equivalent spots as attribute. Once a MSP is generated from multiple RLGS spot patterns, spot pattern differences commonly observed among the constituent RLGS profiles are identified simply by examining the attribute of each master spot in the MSP.

Our approach to generate master spot pattern from multiple RLGS profiles is summarized in Fig. 3. At first, the individual RLGS spot pattern is mapped onto a reference spot pattern, with correcting the non-linear distortion of the gels. To correct the non-linear distortion, we used the equivalent spots detected in the preceding pairwise spot pattern matching. After all of the RLGS spot patterns are mapped onto a reference pattern, the mapped spots are united by means of hierarchical cluster analysis. In the following subsections, we describe both algorithms in detail.

3.1 Delaunay triangular transformation

Here, let us consider the mapping of landmark spots in RLGS profile (a) onto the reference RLGS profile (b). RLGS profiles have asymmetric distortion which differs in position by position on each gel. Because of such irregularity, two RLGS spot patterns cannot be overlaid even if the overall scale and translation factors are adjusted. Accordingly, some kind of non-linear image transformation should be applied to RLGS spot pattern (a) prior to the spot mapping.

We defined such transformation using equivalent spots found on both patterns as anchor points. At first, a DN is constructed from the equivalent spots on the reference pattern (b). Once the DN (b) is constructed, the DN (a) which has same topology can be uniquely defined using the equivalent spots on the spot pattern (a). The RLGS spot pattern (a) is, then, transformed as each node in the DN (a) should completely overlaid on the equivalent node in the reference DN (b).

Any spot inside the Delaunay triangle of DN (a) is linearly transformed into the inside of corre-

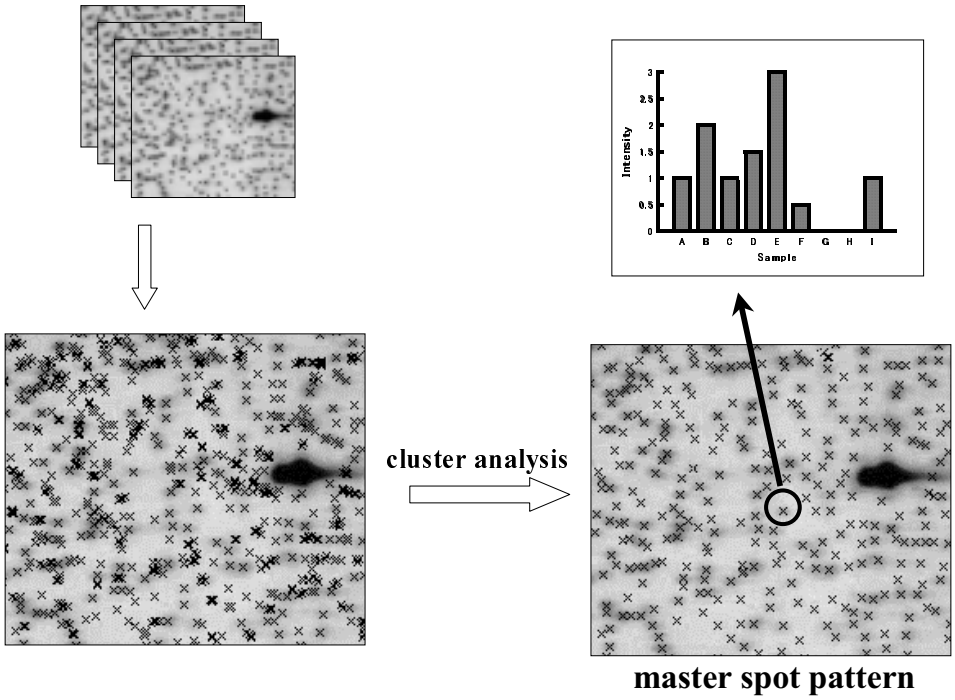
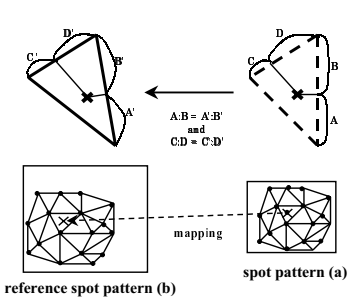
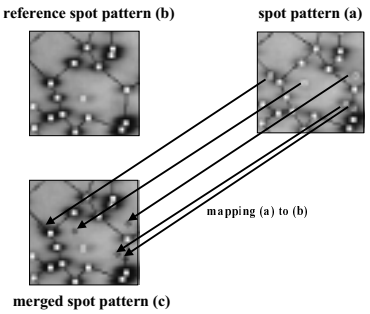


Figure 3: The master spot pattern constructed from several RLGS spot patterns. At first, individual spot patterns are mapped onto a reference pattern. Then, the mapped spots are unified by means of hierarchical cluster analysis (see text). On the master spot pattern, each spot holds intensity information of the equivalent spots in the constituent RLGS profiles. Such information gives much clues to the succeeding genetic analyses.



(a)



(b)

Figure 4: RLGS spot pattern transformation based on Delaunay triangulation. All the spots on the RLGS spot pattern **a** are transformed onto the pattern **b** so as to fit the Delaunay triangles constructed from the equivalent spots.

sponding Delaunay triangle of DN (b) as shown in Fig. 4a. Fig. 4b illustrates the spot mapping using this transformation. In this figure, equivalent spot on the RLGS pattern (a) (depicted as filled square) is mapped to the corresponding spot on the reference pattern (b) (also depicted as filled square), while the spot on pattern (a) which has no equivalent spot on pattern (b) (depicted as open square) is mapped to inside the DN (b) (depicted as open diamond).

With the non-linear transformation explained above, a huge number of RLGS spot patterns can be mapped onto the reference spot pattern, based on the pairwise matching of each RLGS pattern and the reference pattern. As the next step, all of the mapped spots are unified on the reference pattern as described in the next subsection, to give a MSP.

3.2 Automated master spot pattern generation

Let us consider mapping both the RLGS spot patterns (a) and (b) onto a reference pattern (c) as shown in Fig. 5. In this figure, two spots inside the corresponding Delaunay triangles in (a) and (b) are mapped onto the Delaunay triangle in (c). Here both the spots have no equivalent spot on the reference, while they are equivalent each other. In this case, it is quite possible that either of the transformed spots does not occupy the same position on (c), because the mapping (a) to (c) and the mapping (b) to (c) are independent. However, it is also quite possible that each spot is projected to close position. Such indirectly equivalent spots should be unified to give a master spot in a MSP.

We realized the indirectly equivalent spot detection by means of hierarchical cluster analysis, i.e., the mapped spots which occupy the relatively close position on the reference pattern are considered to be equivalent and unified as a cluster. In the hierarchical cluster analysis of the mapped spots, the distance between two clusters is defined as the following equations:

$$D_{ij}^2 = S_{ij}^x + S_{ij}^y,$$

$$S_{ij}^x = \frac{1}{n_{ij}} \sum_{i,j} (x - \bar{x})^2, \qquad S_{ij}^y = \frac{1}{n_{ij}} \sum_{i,j} (y - \bar{y})^2,$$

$$\bar{x} = \frac{1}{n_{ij}} \sum_{i,j} x \quad \text{and} \quad \bar{y} = \frac{1}{n_{ij}} \sum_{i,j} y,$$

where D_{ij} is the distance between cluster i and j, while (x, y) denotes the position of a mapped spot in the reference spot pattern coordinate. Additionally, $\sum_{i,j} a$ instructs the summation of a over the all spots contained in cluster i and j (the number of such spots is denoted as n_{ij}). During the hierarchical clustering of mapped spots, two clusters are not unifed further when the distance between them is larger than the threshold of τ, in order to avoid large cluster formation.

Once the master spot pattern (MSP) is generated from multiple RLGS spot patterns, further analyses can be carried out on the MSP, because a master spot on the MSP takes information about the corresponding spots on the constituent RLGS spot patterns. You can distinguish the specific and statistically significant RLGS pattern difference , separating from the difference caused by the individual or organic difference, simply by examining attributes of each master spots. Additionally, you can also identify RLGS spot translocation which is caused by the mutation occurred at or near to the restriction enzyme recognition sites, by finding master spots which appear or disappear together on a MSP. Such RLGS spot translocation cannot be identified, in principle, by any analysis of matched two RLGS spot patterns.

In the next section, we demonstrate the availability of the MSP , in the hunting of the pathogenic DNA molecular changes.

4 RLGS pattern analysis of colon tumor tissues

In order to demonstrate how well our automated master spot pattern generation algorithm works with a large number of RLGS profiles, and to reveal the availability for hunting pathogenic DNA molecular change, we performed here the analysis of 33 RLGS profiles derived from human colon tissues. These profiles were imaged using *NotI* as the landmark enzyme and using *EcoRV* and *HinfI* as the expansion enzymes. Each RLGS profile was obtained from colon tumor tissue or colon normal tissue sampled from several terminal patients of colon cancer; 15 RLGS profiles of normal tissue (normal profiles), and 18 RLGS profiles of tumor tissue (tumor profiles), as shown in Fig. 6.

At the beginning of the analysis, spot locations and their intensities were recognized by our automated spot recognition algorithm, then we chosen a reference normal profile, on which 2561 landmark spots were recognized successfully. After the remaining 32 RLGS spot patterns were matched to and mapped onto the reference, the master spot pattern (MSP) was obtained by means of hierarchical cluster analysis, as shown in Fig. 7. This MSP contained 3732 master spots, on each of which the information about the landmark spots found on 33 RLGS profiles are mapped. We should notice here

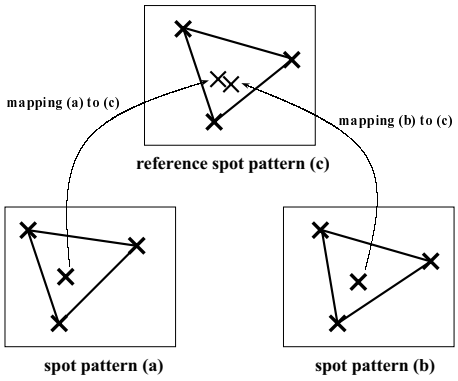
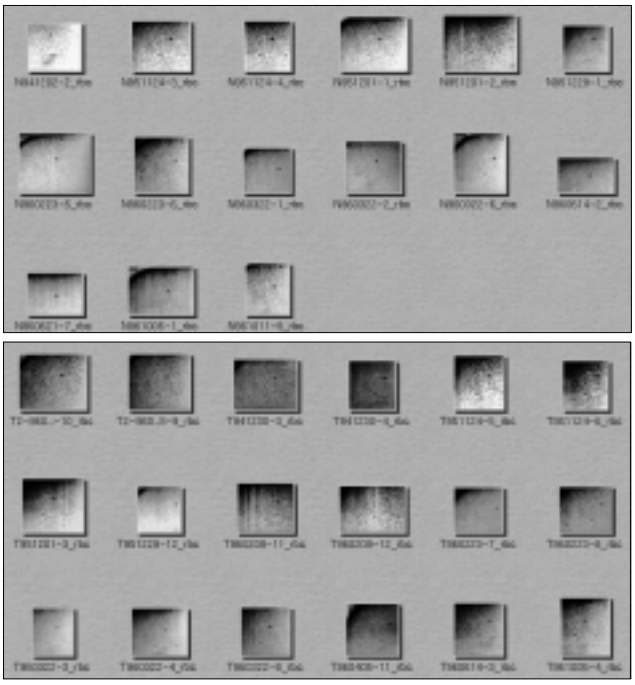


Figure 5: Two spots, one from RLGS pattern (a) and another from RLGS pattern (b), are mapped onto the reference pattern (c). Here, both the spots have no equivalent spot on the pattern (c), while they are equivalent. In this case, either of the transformed spots does not occupy the same position on the reference pattern (c).



(a)

(b)

Figure 6: The RLGS profiles derived from the colon tissues sampled from several terminal patients of colon cancer. (a): 15 RLGS profiles concerning to the normal colon tissue. (b): 18 RLGS profiles derived from colon tumor tissues.

that the MSP was obtained without any human inspection except for choosing the reference.

Now let us distinguish the spot pattern difference which is specific to tumor tissue, separating from the difference caused by the individual difference amongst the patients. When you count the landmark spots which was mapped on each master spot, almost of all RLGS profiles should contribute spots to the master spot commonly observed in both of normal and tumor tissues, while only a few RLGS profiles may contribute spots to the master spot concerned with the individual difference.

When you simply count the number of the spot-contributing tumor and normal profiles separately, the master spots which significantly appear on tumor/normal profiles but disappear on normal/tumor profiles are easily distinguished. In Fig. 8, for instance, we depicted the spot specifically observed in colon tumor profiles. The spot appeared on 15 of 18 colon tumor profiles, while the only one normal profile had the corresponding spot.

The DNA molecular change which lead the above significant RLGS spot pattern difference was not yet characterized. It is quite possible, however, that the DNA methylation status is changed at the corresponding *NotI* locus, since (i) *NotI* is a DNA methylation sensitive restriction enzyme, (ii) methylation status change of some repeat sequence has been found in tumor cells such as neuroblastoma [8, 12]. Further experimental analyses are required to determine the specific DNA molecular change and to elucidate the correlation of such molecular change and tumor. We are planning to specify the loci corresponding to the significantly changed RLGS spots in colon tumor, by means of sequencing of the DNA fragments extracted from the gel, followed by the sequence database analyses.

In the above example, we have shown only the common spot pattern difference identified between the tumor and the normal profiles througth the quite simple analysis of the MSP. It is well-known, however, that the same type of cancer can be caused by different sets of DNA molecular changes. In order to detect such DNA molecular changes, the intensity profiles of the master spots should be analyzed more precisely. As described in the above section, each master spot on a MSP holds intensity information of the equivallent spots in all the constitutent RLGS profiles. When you plot the spot

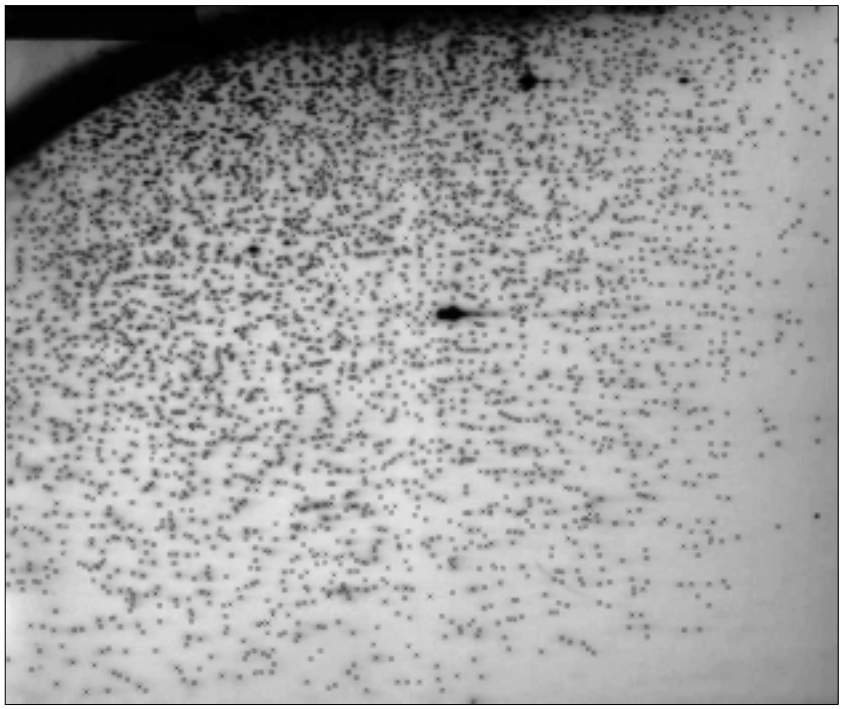


Figure 7: The master spot pattern generated from 33 RLGS profiles of human colon tissues, including 18 tumor tissues. 3732 master spots are drawn overlaid on the reference RLGS profile.

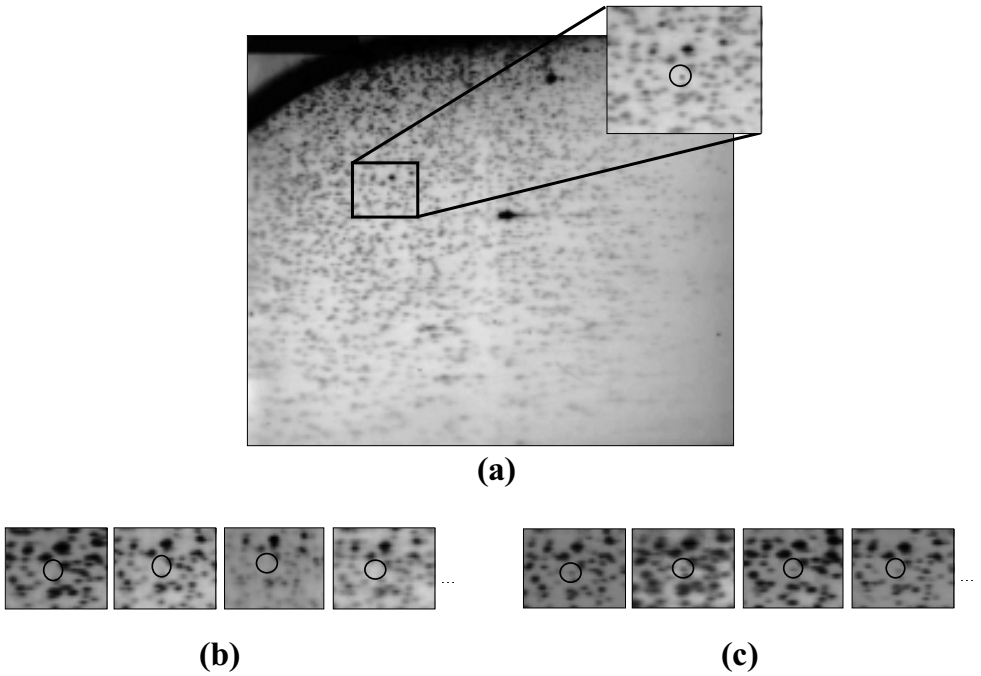


Figure 8: The spot specifically observed in colon tumor. (a): the master spot which was identified to appear frequently on tumor profile but disappear on normal profiles. (b): the landmark spots which disappeared on 14 of 15 colon normal profiles. (c): the landmark spots which appeared on 15 of 18 colon tumor profiles.

intensities of the equivallent spots in a master spot, you have the intensity profile of the master spot, and you can detect DNA molecular changes occurring at the same time on the same genomic DNA, with collecting master spots of which intensity profiles are strongly correlated positively or negatively. With examinning the intensity profile correlation precisely, you can also distinguish the several different sets of DNA molecular changes which cause the same type of tumor independently.

Yet further experimental and computational analyses are necessary, many clues to the unknown pathogenic DNA molecular changes can be obtained, by applying our fully-automated RLGS image processing algorithms to the analysis of huge number of RLGS profiles of diseased genomic DNAs.

5 Summary

We developed and reported previously the automated spot recognition and pairwise spot pattern comparison algorithms, which effectively dealt with disordered RLGS profiles. In this paper, we described the computer algorithm by which huge number of RLGS spot patterns are mapped onto a reference pattern to generate a so-called 'master spot pattern' of the RLGS profiles. With a master spot pattern, the RLGS spot pattern specifically and significantly differ between diseased and non-diseased DNA molecules are easily distinguished from the ones caused by some individual or organic differences.

We applied our automated master spot pattern generation algorithm to the analysis of 33 colon tumor and colon normal profiles, in order to demonstrate its availability of our algorithm in the hunting of pathogenic DNA molecular changes in tumor cells. As the consequence, we successfully identified several RLGS spot pattern difference significant to the genomic DNA of tumor cells.

Our algorithms process the huge number of disordered RLGS profiles accurately but also rapidly, without any human inspection during the analysis. They also give many clues to the succeeding genetic or biochemical analyses which elucidate the relationship between pathogenic DNA molecular changes and diseases.

6 Acknowledgement

We are grateful to Prof. Ueda at Keio University for providing RLGS autoradiogram images, and to Dr. Hirota at NBT Inc. for his technical supports. Finally, we acknowledge Dr. Hayashizaki at The Institute of Physical and Chemical Research (RIKEN) for his kind suggestions on the first stage of this work.

This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas 'Genome Science' from The Ministry of Education, Science, Sports and Culture of Japan.

References

[1] Asakawa, J., Kuich, R., Neel, J.V., Kodaira, M. Satoh, C., and Hanash, S.M., Genetic variation detected by quantitative analysis of end-labeled genomic DNA fragments, *Proc. Natl. Acad. Sci. USA*, 91:9052–9056, 1994.

[2] Bird, A.P., CpG island as gene markers in the vertebrate nucleus, *Trends Genet.*, 3:342–347, 1987.

[3] Brilliant, M.H., Gondo, Y., and Eicher, E.M., Direct molecular identification of the mouse pink-eyed unstable mutation by genome scanning, *Science*, 242:566–569, 1991.

[4] Drmanac, R., Drmanac, S., Strezoska, Z., Paunesku, T., Labat, I., Zeremski, M., Snoddy, J., Funkhouser, W.K., Koop, B., Hood, L., et al., DNA sequence determination by hybridization: a strategy for efficient large-scale sequencing, *Science*, 260:1649–1652, 1993.

[5] Hatada, I., Hayashizaki, Y., Hirotsune, S., Komatsubara, H., and Mukai, T., A genomic scanning method for higher organisms using restriction sites as landmarks, *Proc. Natl. Acad. Sci. U.S.A.*, 88:9523–9527, 1991.

[6] Hayashizaki, Y., Hirotsune, S., Okazaki, Y., Hatada, I., Shibata, H., Kawai, J., Hirose, K., Watanabe, S., Fushiki, S., Wada, S., et al., Restriction landmark genomic scanning method and its various applications, *Electrophoresis*, 14:251–258, 1993.

[7] Hayashizaki, Y., Hirotsune, S., Okazaki, Y., Shibata, H., Akasako., A., Muramatsu, M., Kawai, J., Hirasawa, T., Watanabe, S., Shiroishi, T., et al., A genetic linkage map of the mouse using Restriction Landmark Genomic Scanning (RLGS), *Genetics*, 138:1207–1238, 1994.

[8] Miwa, M., Yashima, K., Sekine, T., and Sekiya, T., Demethylation of a repetitive DNA sequence in human cancers, *Electrophoresis*, 16:227–232, 1995.

[9] Southern, E.M., Case-Green, S.C., Elder, J.K., Johnson, M., Mir, K.U., Wang, L., and Williams, J.C., Arrays of complementary oligonucleotides for analyzing the hybridization behavior of nucleic acids, *Nucl. Acids Res.*, 22:1368–1373, 1994.

[10] Takahashi, K., Nakazawa, M., and Watanabe, Y., DNAinsight: an image processing system for 2-D gel electrophoresis of genomic DNA," *Genome Informatics*, 8:135–146, 1997.

[11] Takahashi, K., Nakazawa, M., Watanabe, Y., and Konagaya, A., Fully-automated spot recognition and matching algorithms for 2-D gel electrophoretogram of genomic DNA, *Genome Informatics*, 9:161–172, 1998.

[12] Thoraval, D., Asakawa, J., Wimmer, K., Kuick, R., Lamb, B., Richardson, B., Ambros, P., Glover, T., and Hanash, S., Demethylation of repetitive DNA sequences in neuroblastoma. *Genes, Chromosomes and Cancer*, 12:234–244, 1996.

[13] Toussaint, G.T., The relative neighborhood graph of finite planar set, *Pattern Recognition*, 12:261–268, 1980.

[14] "To affinity ... and beyond!," *Nature Genet.*, 14:367-370, 1996.

[15] Watanabe, Y., Takahashi, K., and Nakazawa, M., Automated detection and matching of spots in autoradiogram images of two-dimensional electrophoresis for high-speed genome scanning, *Proc. International Conference on Image Processing, IEEE Signal Processing Society*, III:496-499, 1997.

[16] Yershov, G., Barskey, V., Belgovskiy, A., Kirillov, E. Kreindlin, E., Ivanov, I., Parinov, S., Guschin, D., Drobishev, A., Dubiley, S., and Mirzabekov, A., DNA analysis and diagnostics on oligonucleotide microchips, *Proc. Natl. Acad. Sci. USA*, 93:4913–4918, 1996.