

國立交通大學

統計學研究所

碩士論文

利用混合模型對微陣列影像做影像切割

Segmentation of Microarray Images  
By Mixture Models



研究生：藍秀仁

指導教授：盧鴻興博士

中華民國 九十四 年 六 月

利用混合模型對微陣列影像做影像切割

Segmentation of Microarray Images  
By Mixture Models

研究生：藍秀仁                      Student : Hsiu-Jen Lan  
指導教授：盧鴻興博士              Advisor : H. H.-S. Lu

國立交通大學理學院

統計研究所

碩士論文



A thesis

Submitted to Institute of Statistics  
College of Science  
National Chiao Tung University  
in partial Fulfillment of the Requirements  
For the Degree of  
Master  
in  
Statistics

June 2005

Hsinchu, Taiwan, Republic of China

中華民國 九十四 年 六 月

## 誌謝

本論文能夠順利的完成，首先要感謝我的指導老師盧鴻興老師，感謝老師在這一年多來對我不厭其煩地教導，使我學習了真正做研究的方法和態度，另外也要感謝陳泰賓學長這一年多對我的指導，在學長的身上學到了凡事只要肯努力、一定能成功的做人態度。當然最重要的是感謝我的父母，謝謝你們給予我精神上和經濟上的支持，使我能夠順利的完成學業。在日常生活中，感謝雅靜及所有408研究室的同學們，在我遇到挫折失敗時，能給我許多鼓勵並陪我渡過快樂的每一天。最後特別感謝我的論文的口試委員徐南蓉老師、黃冠華老師和許文郁老師，於百忙之中撥冗審稿、校正，並在口試時給我的諸多指導。



藍秀仁 謹誌于  
國立交通大學統計研究所  
中華民國九十四年六月

# 利用混合模型對微陣列影像做影像切割

研究生：藍秀仁

指導教授：盧鴻興博士

國立交通大學統計研究所

## 中文摘要

微陣列技術是利用 cDNA 或寡核苷酸(oligonucleotide)的探針進行雜交反應(hybridization)去偵測 RNA 在樣本內的表現水準。藉由雷射掃描器可檢測探針與目標染料的雜交對所放出的螢光量。因此，一台掃描器發現的螢光量與 RNA 的表現水準有關。所以微陣列影像的強度分析對於測量 RNA 的表現水準是重要的步驟。

在此研究中，我們將考慮雙色微陣列，如紅色或綠色影像即是分別用 Cy5 或 Cy3 染劑於控制組與對照組樣本上。具體而言，我們將應用混合模型去分割出訊號值及背景值。這是因為混合模型可以靈活地使用不同的參數，例如位置及尺度參數，去建構訊號值及背景值的分配。

我們透過微陣列影像上已知 Cy5 和 Cy3 目標比率的 spike genes，來比較混合模型與 GenePix 6.0 軟體的切割成效。計算 spike genes 的平均測量比率與標地比率之間的相對均誤差來衡量影像切割的成效。我們在本研究中發現，由混合模型的切割方法可以降低其相對均誤差及改善測量比率的準確性。

## Abstract

Microarray technique uses the hybridization of complementary DNA or oligonucleotide segments of probes and targets to detect the expression level of RNAs in samples. The hybridized pairs of probes and targets with dyes will emit fluorescence during the scanning process of a laser scanner. Thus, the fluorescence detected by a laser scanner is related to the expression level of RNAs in samples. Hence, it is important to analysis the intensities of microarray images to measure the detected fluorescence for expression levels of RNAs.

In this study, we will consider the image analysis of two-channel cDNA microarrays that have red and green images for Cy5 and Cy3 dyes in control and experiment samples. Specifically, we will apply mixture models to segment the foreground and background images because the mixture models are flexible to model the distributions of foreground and background intensities with different parameters that include location and scale parameters.

We will compare the performance of segmentations by mixture models with those by the software of GenePix 6.0 through microarray images with spike genes that have target ratios between Cy5 and Cy3. Sums of square of relative errors between average estimated ratios and target ratios of spike genes are used to evaluate the performance of segmentation results. The segmentation methods by mixture models can reduce the sums of square of relative errors and improve the accuracy of estimated ratios.

## Contents:

<b>Chapter 1.</b>	<b>Introduction</b>	<b>1</b>
<b>Chapter 2.</b>	<b>Steps of Image Processing</b>	<b>3</b>
2.1	Overview	3
2.2	Operation Steps in GenePix 6.0	4
2.3	Performances of Segmentation in GenePix 6.0	4
<b>Chapter 3.</b>	<b>Image Segmentation by Mixture Models</b>	<b>6</b>
3.1	Flowchart	6
3.2	Combined Images	7
3.3	EM Algorithm for a Normal Mixture Model (NMM)	8
3.4	Gaussian Smoothing	12
3.5	Normalization Factors	12
<b>Chapter 4.</b>	<b>Empirical Studies</b>	<b>13</b>
4.1	Data and Statistics	14
4.2	Segmentation Results by GenePix 6.0 and NMM	15
4.3	Comparisons of Results by GenePix 6.0 and NMM	18
<b>Chapter 5.</b>	<b>Conclusion and Discussion</b>	<b>22</b>
	<b>References</b>	<b>23</b>
	<b>Appendix</b>	<b>26</b>