

ห้องสมุดงานวิจัย สำนักงานคณะกรรมการวิจัยแห่งชาติ



E47216

**IDENTIFICATION OF RECTAL CANCER GENES  
BY MICROARRAY ANALYSIS**

**NIPAPORN THIPMANEE**

**MASTER OF SCIENCE  
IN BIOINFORMATICS**

**THE GRADUATE SCHOOL  
CHIANG MAI UNIVERSITY  
SEPTEMBER 2011**

๐๐๐ 254 128

ห้องสมุดงานวิจัย สำนักงานคณะกรรมการการวิจัยแห่งชาติ



E47216

**IDENTIFICATION OF RECTAL CANCER GENES  
BY MICROARRAY ANALYSIS**



**NIPAPORN THIPMANEE**

**A THESIS SUBMITTED TO THE GRADUATE SCHOOL IN  
PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF  
MASTER OF SCIENCE  
IN BIOINFORMATICS**

**THE GRADUATE SCHOOL  
CHIANG MAI UNIVERSITY  
SEPTEMBER 2011**

**IDENTIFICATION OF RECTAL CANCER GENES  
BY MICROARRAY ANALYSIS**

NIPAPORN THIPMANEE

THIS THESIS HAS BEEN APPROVED  
TO BE A PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF MASTER OF SCIENCE  
IN BIOINFORMATICS

**EXAMINING COMMITTEE**

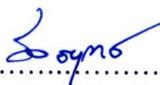
  
.....CHAIRPERSON  
Asst. Prof. Dr. Lily Ingrisawang

  
.....MEMBER  
Assoc. Prof. Dr. Jeerayut Chaijaruwanich

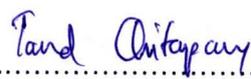
  
.....MEMBER  
Assoc. Prof. Imjai Chitapanarux, MD

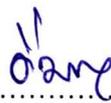
**THESIS ADVISORY COMMITTEE**

  
..... ADVISOR  
Asst. Prof. Dr. Sukon Prasitwattanaseree

  
..... CO-ADVISOR  
Assoc. Prof. Dr. Jeerayut Chaijaruwanich

  
..... CO-ADVISOR  
Dr. Nisa Chawapun

  
..... CO-ADVISOR  
Asst. Prof. Taned Chitapanarux, MD

  
..... CO-ADVISOR  
Assoc. Prof. Imjai Chitapanarux, MD

## ACKNOWLEDGEMENTS

Many people have helped me throughout my degree. I feel proud of my decision to study Bioinformatics and do this research.

I am grateful to acknowledge Dr. Sukon Prasitwattanaseree for very important suggestions on my work during my graduate study at Chiang Mai University, in my work. Although, he has many duties, he spends his time to provide valuable guidance to me.

I wish to acknowledge my advisor Dr. Jeerayut Chaijaruwanich, Dr. Nisa Chawapun and Dr. Imjai Chitapanarux, MD. During a period of my graduate student, I am truly grateful to Dr. Jeerayut, Dr. Nisa and Dr. Imjai for their patience and guidance throughout this research. Especially, I would like to express my deeply grateful to Dr. Nisa for her generous support and inspiration for me. I would also like to acknowledge Dr. Taned Chitapanarux, MD for serving as my thesis co-advisor.

I would also like to acknowledge Dr. Lily Ingsrisawang for being on my examining committee.

I am very grateful for all members of Bioinformatics Research Laboratory (BIRL) and Cancer Treatment and Research Center. I am greatly indebted to Mr. Keng Jiamkitwattana for conducting microarray raw data for my thesis and supporting me everything.

Special thank to Science Achievement Scholarship of Thailand for providing financial support.

My utmost gratitude, love, support and encouragement, of course, of my family, I am truthfully grateful to my parents, sister and brother for everything. I dedicate this thesis to all my beloved people.

Nipaporn Thipmanee

<b>Thesis Title</b>	Identification of Rectal Cancer Genes by Microarray Analysis	
<b>Author</b>	Ms. Nipaporn Thipmanee	
<b>Degree</b>	Master of Science (Bioinformatics)	
<b>Thesis Advisory Committee</b>	Asst. Prof. Dr. Sukon Prasitwattanaseree	Advisor
	Assoc. Prof. Dr. Jeerayut Chaijaruwanich	Co-advisor
	Dr. Nisa Chawapun	Co-advisor
	Asst. Prof. Taned Chitapanarux, M.D.	Co-advisor
	Assoc. Prof. Imjai Chitapanarux, M.D.	Co-advisor

## ABSTRACT

**E47216**

DNA Microarray is a powerful technology to study in genomic filed, especially to analyze a large number gene expression in a single experiment. This study concentrated on rectal carcinogenesis and aimed to identify the profiles of differentially expressed genes by comparing normal and rectal cancer tissues using cDNA microarray. The gene expression profiles were constructed from normal and tumor tissues in 6 stage III rectal cancer patients treated at Maharaj Nakorn Chiang Mai Hospital between December 2009 – April 2010. cDNA microarray analysis for rectal cancer genes was

**E47216**

completed using the R program and linear models for microarray data or limma package for analysis gene expression microarray data could clearly distinguish the gene profiles of rectal cancer from those of gene from normal tissues as p-value of t-test less than 0.005. Moreover, the status of gene expression defined as 2-fold. Finally, cDNA microarray identified 12 genes differentially expressed between cancer and normal tissues. Among 12 genes, 8 genes were down-regulated and 4 genes were up-regulated in rectal cancer.

ชื่อเรื่องวิทยานิพนธ์	การระบุยีนมะเร็งลำไส้ตรงโดยการวิเคราะห์ไมโครอะเรย์	
ผู้เขียน	นางสาว นิภาพร ทิพย์มะณี	
ปริญญา	วิทยาศาสตรมหาบัณฑิต (ชีวสารสนเทศศาสตร์)	
คณะกรรมการที่ปรึกษาวิทยานิพนธ์	ผศ.ดร.สุคนธ์ ประสิทธิ์วัฒน์เสรี	อาจารย์ที่ปรึกษาหลัก
	รศ.ดร.จิรยุทธ ไชยจรรูมิช	อาจารย์ที่ปรึกษาร่วม
	ดร.นิสา ชวพันธุ์	อาจารย์ที่ปรึกษาร่วม
	ผศ.นพ.ธเนศ ชิตาพนารักษ์	อาจารย์ที่ปรึกษาร่วม
	รศ.พญ.อิมใจ ชิตาพนารักษ์	อาจารย์ที่ปรึกษาร่วม

### บทคัดย่อ

**E47216**

DNA ไมโครอะเรย์เป็นเทคโนโลยีที่มีประสิทธิภาพสำหรับการศึกษาทางด้านจีโนมิกส์ ไมโครอะเรย์ถูกใช้ในการวิเคราะห์ข้อมูลการแสดงออกของยีนจำนวนมาก ในการศึกษาวิจัยนี้สนใจ ยีนที่ทำให้เกิดมะเร็งลำไส้ตรง วัตถุประสงค์ของการศึกษานี้คือใช้ cDNA ไมโครอะเรย์ในการระบุยีน ที่มีค่าการแสดงออกที่แตกต่างกันระหว่างเนื้อเยื่อปกติและเนื้อเยื่อมะเร็งในผู้ป่วยมะเร็งลำไส้ตรงระยะ III จำนวน 6 คนซึ่งเข้ารับการรักษาที่โรงพยาบาลมหาราชนครเชียงใหม่ระหว่างเดือนธันวาคมปี 2552 ถึงเดือนเมษายนปี 2553 การวิเคราะห์ไมโครอะเรย์ของยีนมะเร็งลำไส้ตรงได้ใช้โปรแกรม R การวัด การแสดงออกของยีนโดยใช้แบบจำลองเชิงเส้นแยกแยะการแสดงออกของยีนกลุ่มมะเร็งลำไส้ตรง ออกจากกลุ่มยีนปกติ โดยใช้การทดสอบทางสถิติ แบบ t-test ที่ p-value น้อยกว่า 0.005 และค่า fold change เท่ากับ 2 ในการพิจารณารูปแบบการแสดงออกของยีนมะเร็ง ผลการศึกษาวิจัยนี้พบ 12 ยีนที่มี ค่าการแสดงออกที่แตกต่างระหว่างชิ้นเนื้อมะเร็งและชิ้นเนื้อปกติ ซึ่ง 8 ยีนมีการแสดงออกแบบ down-regulated และ 4 ยีนที่เหลือมีการแสดงออกแบบ up-regulated ในมะเร็งลำไส้ตรง

**TABLE OF CONTENTS**

	<b>Page</b>
<b>ACKNOWLEDGMENTS</b>	iii
<b>ABSTRACT (ENGLISH)</b>	v
<b>ABSTRACT (THAI)</b>	vii
<b>LIST OF TABLES</b>	xi
<b>LIST OF FIGURES</b>	xii
<b>ABBREVIATIONS</b>	xiv
<b>CHAPTER I INTRODUCTION</b>	
1.1 Research objectives	4
1.2 Usefulness of the research (theoretical and/or applied)	5
1.3 Research scope	5
1.4 Research location	6
1.5 Research duration	6
<b>CHAPTER II cDNA MICROARRAY</b>	
2.1 History of microarray	7
2.2 Definition of microarray	8
2.3 Type of microarray	10
2.4 The procedure to analyze microarray data	11
2.4.1 Biological question	12
2.4.2 Experiment design	15
2.4.3 Microarray experiment	17
2.4.4 Pre-processing data	19

2.4.4.1	Image analysis	20
2.4.4.2	Expression quantification	26
2.4.4.3	Normalization	34
2.4.5	Statistical analysis	38
<b>CHAPTER III COLORECTAL CANCER</b>		
3.1	Epidemiology	44
3.2	Biologic characteristics and pathology	45
3.3	Type of colorectal cancer	47
3.4	Cancer staging	48
<b>CHAPTER IV METHOD TO IDENTIFY RECTAL CANCER GENES WITH SIGNIFICANT DIFFERENTIAL EXPRESSION BY MICROARRAY ANALYSIS</b>		
4.1	Introduction	52
4.2	Tissue sample	52
4.3	Gene expression study by cDNA microarray	54
4.4	Microarray data analysis	57
<b>CHAPTER V RESULTS</b>		
5.1	Data before pre-processing	67
5.2	Data pre-processing	72
5.3	Gene differential expression	78
<b>CHAPTER VI DISCUSSION AND CONCLUSION</b>		
6.1	Discussion	82
6.2	Conclusion	91

<b>REFERENCES</b>	93
<b>APPENDICES</b>	108
<b>APPENDIX A</b>	109
<b>APPENDIX B</b>	114
<b>APPENDIX C</b>	136
<b>CURRICULUM VITAE</b>	137

## LIST OF TABLES

<b>Table</b>	<b>Page</b>
2.1 Segmentation methods and examples of software implementation	25
3.1 TNM staging definition for colorectal cancer	50
4.1 Six patients of rectal cancer whose contain tumor of rectal in stage III	53
4.2 Some section of the output file of ScanArray Express Software	56
4.3 The layout of targets file for reading six arrays into R software	58
4.4 The information of ten genes in the top of Human OpArray.GAL	58
4.5 Eight spot types of microarray data	59
4.6 The targets frame of factorial design	62
4.7 Some sections of the information provided by PerkinElmer Inc in USA	64
4.8 Some information of GO database	66
5.1 Ten values of foreground and background intensities of red and green channels	69
5.2 Ten values of red and green intensities after normexp background correction	72
5.3 M and A values of normalization method	74
5.4 Twelve genes differentially expressed less than p-value at 0.005, comparing normal and cancer cells from the pooled data set	79
6.1 The statistical values of six difference background methods	83
6.2 The statistical values of three difference within-array methods	85
B.1: Two hundred and fifty-six genes differentially expressed less than p-value at 0.005, comparing normal and cancer cells from the pooled data set	114

**LIST OF FIGURES**

<b>Figure</b>	<b>Page</b>
2.1 Anatomy of DNA microarray	9
2.2 Target and probe	9
2.3 The six steps of microarray analysis cycle	11
2.4 Bioinformatics concept	13
2.5 Central dogma	14
2.6 Commonly design of microarray	16
2.7 The procedure of microarray experiment	18
2.8 The ideal of image processing	21
2.9 Grid array patch on the array	22
2.10 Overview spots on microarray glass slide	23
2.11 Four segmentation method of image processing	24
2.12 The various intensity distribution of whole microarray spots	27
2.13 3D background correction	28
2.14 The concept of normexp background correction method	30
2.15 Adaptive background correction	32
2.16 Anatomy of box plot	33
2.17 The concept of MA-plot	34
2.18 Locally weight polynomial regression (Loess regression)	37
2.19 MA-plot of before and after normalization by loess method	37
3.1 The diversity of cancer	44

3.2 The anatomy of colon and rectum	45
3.3 Genetic progression of colorectal cancer	46
3.4 Colorectal staging	49
4.1 The reference design of thesis data	54
5.1 The visualization of six raw microarray data	68
5.2 MA-plot of six raw microarray data	71
5.3 The visualization of M and A values	75
5.4 The density plot of six raw microarray data	76
5.5 The density plot of six normalization data	77
6.1 MA-plot of six difference background correction methods	84
6.2 MA-plot of three difference within-array normalization method	85
6.3 Density plot of three difference within-array normalization methods	86
6.4 MA-plot and density plot of Aquantile between-array method	87

## ABBREVIATIONS

<b>AJCC</b>	American Joint Committee On Cancer
<b>TNM</b>	T= primary tumor, N= lymph node, M= metastasis
<b>WD</b>	well-differentiated
<b>PD</b>	poorly differentiated
<b>DNA</b>	deoxyribonucleic acid
<b>Limma</b>	linear models for microarray data
<b>RNA</b>	ribonucleic acid
<b>Lowess</b>	locally weighted scatterplot smoothing
<b>Aquantile</b>	applying quantile normalization to the A-values
<b>Cy5</b>	red-fluorescing cyanine dyes
<b>Cy3</b>	green-fluorescing cyanine dyes
<b>cDNA s</b>	complementary DNA
<b>mRNA</b>	messenger RNA
<b>SNP</b>	Single Nucleotide Polymorphism
<b>PCR</b>	the polymerase chain reaction
<b>aRNA</b>	poly [A] RNA
<b>normexp</b>	normal-exponential convolution model

<b>IQR</b>	inter-quartile rang
<b>CRC</b>	colorectal cancer
<b>APC</b>	adenomatous polyposis coli.
<b>HNPCC</b>	hereditary nonpolyposis colorectal cancer
<b>FAP</b>	familial adenomatous polyposis
<b>CT</b>	X-ray computed tomography
<b>MRI</b>	magnetic resonance imaging
<b>1st Qu</b>	first quartile
<b>3rd Qu</b>	third quartile
<b>FDR</b>	false discovery rate