CHAPTER VI

DISCUSSION AND CONCLUSION

6.1 Discussion

This study, we identified genes differentially expressed between normal and rectal cancer tissues from two color microarray platforms using the statistical algorithm provided for R software. 6 stage III rectal cancer patients were enrolled and a total of 12 samples, six normal and six tumor samples were analyzed. In the process of quality assessment by comparisons the level of gene expression across six arrays, we found that the quality of the first array was not as good as the others. This might be due to the incorporation of the fluorescent molecules into mRNA molecule at the time of conversion to cDNA was not so efficient. However, we performed normalization for all six arrays. In this research, we reviewed various background correction methods of two-color microarray data and subsequently selected six background correction methods to test with our data. Figure 6.1 shows the MA-plot of the differences between background correction methods that analyzed with the same raw data. The statistical values of each method are also presented. Among six methods, the normexp was most apparent at low M- and A-values and MA-plot in Figure 6.1 supported for low variance of data analyzed by normexp method. Therefore, we selected the normexp background correction method for our data analysis.

Table 6.1 The statistical values of six difference background methods.

Method		Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
none	M	-5.812	-3.712	-3.384	-3.256	-2.979	1.785
	A	8.883	10.500	10.690	10.670	10.860	15.460
substract	M	-Inf	-4	-4	-	-3	Inf
	A	-Inf	9	10	-Inf	10	15
minimum	M	-14.980	-4.519	-3.911	-3.929	-3.366	10.930
	A	3.254	9.296	9.757	9.473	10.040	15.440
movingmin	M	-6.163	-4.053	-3.709	-3.567	-3.280	1.800
	A	8.43	10.28	10.48	10.46	10.66	15.45
normexp	M	-8.553	-3.745	-3.050	-3.051	-2.376	3.913
	A	6.610	7.919	8.277	8.315	8.625	15.390
edward	M	-11.450	-4.519	-3.911	-3.883	-3.366	6.452
	A	5.507	9.296	9.757	9.542	10.040	15.440

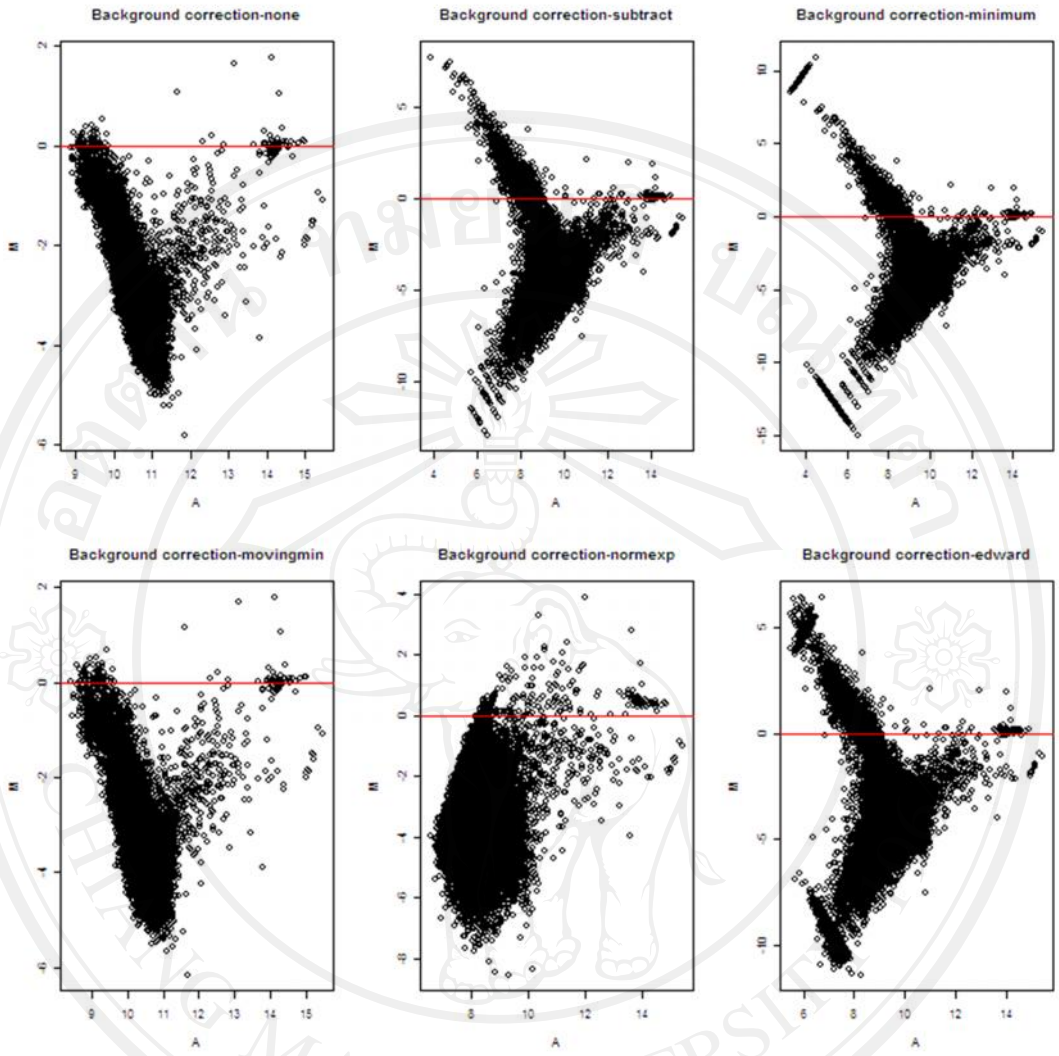

Figure 6.1 MA-plot of six difference background correction methods.

For the within-array normalization, we compared three within-array methods for analysis gene expression values of normexp background correction. The results of three methods are shown with MA-plot and density plot in Figure 6.2 and Figure 6.3, respectively. The statistical values of M and A-values for each within-array normalization method are shown in Table 6.2.

Table 6.2 The statistical values of three difference within-array methods.

Method		Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
loess	M	-9.43385	-0.99669	0.03025	-0.03904	0.97647	7.10818
	A	1.639	4.462	4.976	5.018	5.528	14.844
none	M	-10.639	-4.035	-2.949	-2.940	-1.835	4.566
	A	1.639	4.462	4.976	5.018	5.528	14.844
median	M	-7.689259	-1.085381	0.00000	0.009227	1.114630	7.515153
	A	1.639	4.462	4.976	5.018	5.528	14.844

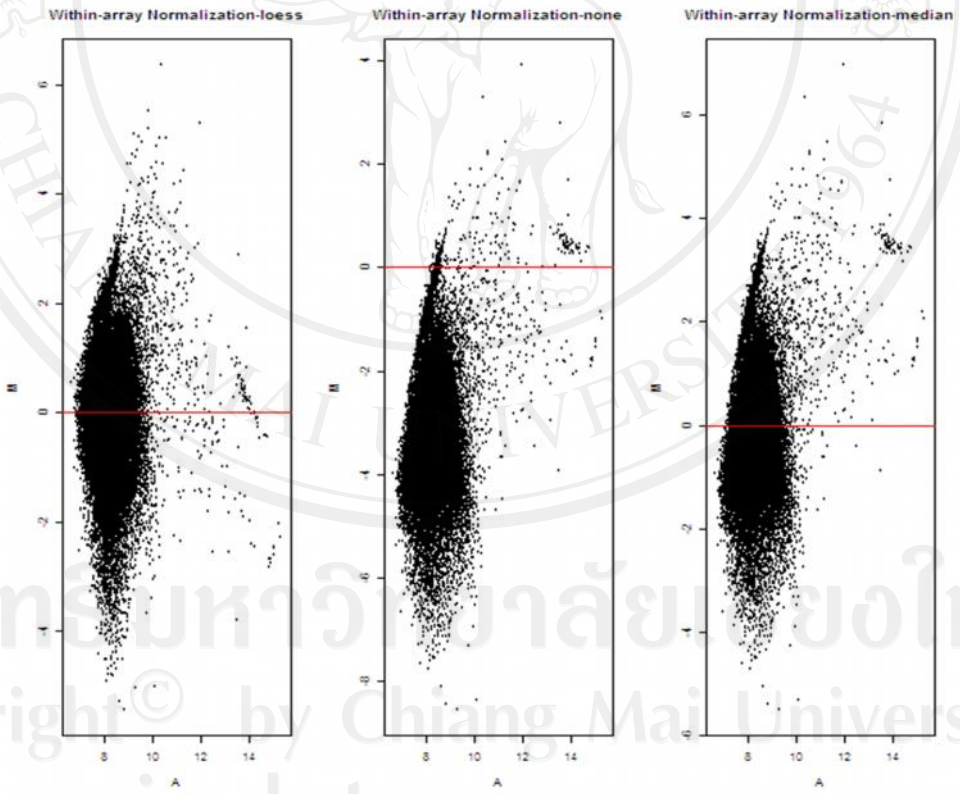

Figure 6.2 MA-plot of three difference within-array normalization methods.

Within-array normalization concentrated with M-value. Thus we mainly considered the MA-plot. However, the densities plot helped assess for the method with good fitting of data. Among three within-array normalization methods, the results of no normalization demonstrated the worst in statistical values and density plot was not good fit while median method was the best. Although, it was shown that median was the best normalization method, the loess method was used in this study since it is used for quantitative analysis in ScanAraray Express software.

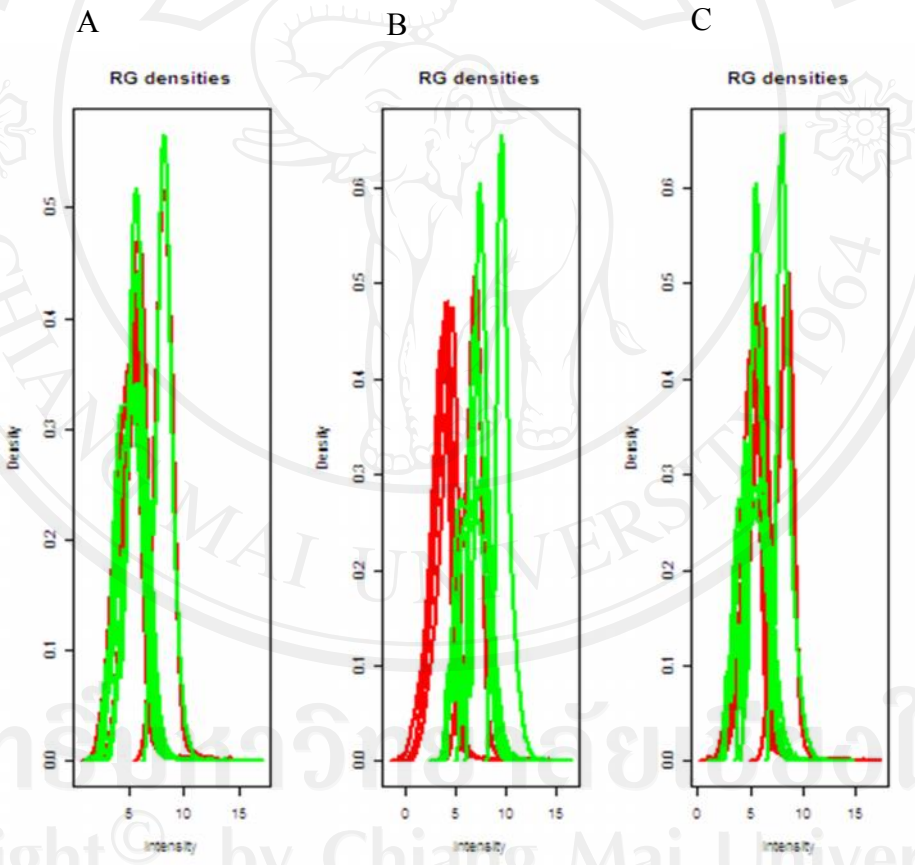

Figure 6.3 Density plot of three difference within-array methods.
A) Density plot of loess method B) Density plot of non method and
C) Density plot of median method.

For between-array normalization, three difference methods were compared including none, scale and Aquantile method. We chose the Aquantile method for normalizing our data. While the within-array normalizing only concerned with M-value and ignore A-value, Aquantile between-array method only analyzed A-value based on quantile method. The MA-plot and density plot in Figure 6.4 shows that Aquantile was the proper method good fitting for our data.

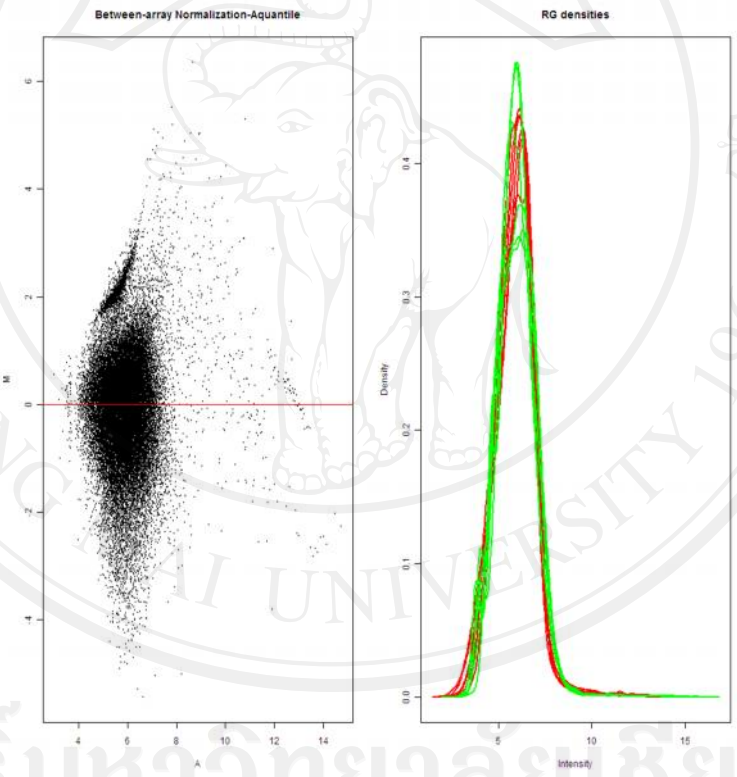

Figure 6.4 MA-plot and density plot of Aquantile method.

We further identified genes profile with differentially expressed between comparison groups by using t-test. The two groups were independently observed (normal vs. tumor tissues). The normalized data were used in between-array normalization statistics for differential expression step. While 2-fold change is the

common method for identification molecular changes between two difference samples.

We found 12 genes significant differentially expressed between normal and rectal cancer tissues. Among 12 genes, 8 genes were down-regulated and 4 genes were up-regulated in rectal cancer (8 down-regulated: CA3, CCT7, DKK3, E3F7, KRT23, LARGE, PTGDS, TOMM34 and 4 up-regulated: HPGD, ID4, RPL10, and RPS2).

Carbonic anhydrase (CA) is a family of enzymes which interconvert carbon dioxide and water to bicarbonate and protons. The primary function of the enzyme is maintainance acid-base balance in blood and other tissues. The member of this enzyme family; CA9 is well-known to be associated with hypoxic tumor by alterations in microenvironmental pH (54). CA3 may have a role in intracellular signaling, particularly in response to oxidative stress. A high level of free radicals in CA3 down-expression cells may affect growth signaling pathways leading to transform normal cells to cancerous cells (55). According to the published database from European Bioinformatics Institute (EBI) (http://www.ebi.ac.uk), CA3 is down regulated in many types of cancer such as breast cancer, bladder cancer, colorectal cancer, lung cancer and colon cancer (56). In our study CA3 was also down-regulated.

CCT7 is a member of the chaperonin containing TCP1 complex (CCT), a molecular chaperone that assists in the folding of actin, tubulin, and other cytosolic proteins. Although correlation between this protein and carcinogenesis has not yet been reported but the closely related protein CCT1 was found over-expressed in human hepatocellular and colonic carcinoma (57) while CCT6 over-expressed in

colonic carcinoma (58). CCT7 was down-regulated in human lymphoblastoid cells, TK6 cell (59). In our study CCT7 was down-regulated.

Dickkopf-related protein 3 (Dkk3) is secreted Wnt antagonist and its functional loss can contribute to activation of the Wnt pathway result in carcinogenesis through dysregulation of cell proliferation and differentiation. The expression of this gene is decreased in cancer of various tissues such as liver (60), breast (61), GI tract (62), and lung (63). We found Dkk3 was under-expressed in rectal cancer that correlated with previous cancer researches.

The E2F family of transcription factors plays a crucial role in regulation of genes involved in cell proliferation, differentiation, and apoptosis de Bruin A *et al*., suggested that E2F7 might function as a tumor suppressor gene and provide an important contribution to the regulation of E2F activities in multiple tissues during development and/or tumorigenesis (64). E2F7 is over-expressed in cutaneous squamous cell carcinomas (65) and lung cancer (66) while down-regulated in ovarian cancer (67). We also found under-expressed of this gene in rectal cancer.

Keratin is a family of fibrous structural proteins. In cancer, many types of keratin are extensively used as diagnostic tumor markers which show both over and under-expression in various types of cancer (68). Keratin23 (KRT23) was found down-regulated in this study.

LARGE is one of the largest in the human genome, encodes a member of the N-acetylglucosaminyltransferase gene family. Mutations in this gene may be involved in the development and progression of meningioma through modification of ganglioside composition and other glycosylated molecules in tumor cells (69).

However the relation between tumor progression and expression level of this gene has not yet been studied. It was under-expressed in our study.

Prostaglandin D2 synthase (PTGDS) catalyzes the conversion of prostaglandin H2 (PGH2) to prostaglandin D2 (PGD2). Various studies showed that this gene is down regulated in brain tumor (70) lung cancer (71) uterine leiomyoma (72), oral cancer (73). It is noticeable that PTGDS gene associated with prostaglandin metabolism possibly that prostaglandin may involve in tumor suppressor function. This may lead to discovery of the novel method to diminish cancer.

Mitochondrial import receptor subunit TOM34 (TOMM34) is involved in the import of precursor proteins into mitochondria. TOMM34 is up-regulated in human colon cancer (74). Various studies in the database of European Bioinformatics Institute (EBI) (http://www.ebi.ac.uk) showed that TOMM34 is down regulated in breast cancer, lung cancer, liver cancer, hereditary leiomyomatosis and renal cell cancer (75). In our study TOMM34 was down-regulated.

15- hydroxyprostaglandin dehydrogenase (HPGD) is a normal physiologic COX-2 inhibitor which catalyzes the conversion of the prostaglandin 15-OH to a 15-keto group (76). This enzyme is considered as tumor suppressor gene in colon cancer (76) gastric cancer (77) pancreatic tumor (78) lung cancer (79) and breast cancer (80). It was up-regulated in prostate cancer (81). However it was over-expressed in our study.

DNA-binding protein inhibitor-4 (ID4) regulates expression of tissue-specific genes. A range of studies showed that ID4 modulation associated with various types of cancer. In cancer of bladder, brain and lung this gene is up-regulated (82-84) while it is down-regulated in cancer of gastric, colorectal, leukemia and lymphoma (85-91).

However, it can be both up- and down-regulated in breast cancer (92-96). In our study this gene was over-expressed.

Up regulation of RPL10, RPS2 was correlated with tumorigenesis in previous studies. 60S ribosomal protein L10 (RPL10) is over-expressed in prostate cancer (97) while 40S ribosomal protein S2 (RPS2) is over-expressed in prostate cancer (98) and liver cancer (99). We notice that both genes may be associated with protein synthesis since their products are ribosomal component.

6.2 Conclusion

In this research study, we focused on the implementation of free software as the computational tool for microarray data analysis without using commercial software. We selected R programming and limma package available for analysis microarray. The objective was to analyze the difference in gene expression between normal and rectal cancer tissues by cDNA microarray. The genes profiles were identified by p-value analysis using *t* tests with cutoff at 0.005 and subsequently used 2-fold change for identified the status of significantly differential of genes expression.

Among 12 genes obtained, we draw our attention to the five genes that might help improve the efficiency of cancer management. HPGD and PTGDS are the most important for colorectal carcinogenesis as they are involving the tumor suppressor function. RPL10 and RPS2 are associated with many types of cancer and these genes associated with the development of colon cancer. Last but not least, TOMM 34 is more interesting. The previous research used TOMM 34 to study gene expression of colon cancer, alteration with a novel drug target and TOMM34 was effectively suppressed and drastically inhibited cell growth (74). It is fascinating that TOMM34

associated translocase of outer mitochondrial membrane 34. How TOMM34 gene involves in colorectal cancer, how mitochondrial protein associated with cancer, which biological pathway associates with the regulation of TOMM34, these are interesting question waiting for explore. Using real time PCR for validation of five genes of microarray resulting analyzing by R programming is the challenge of future work.

The framework of this research study has found to be advantageous for three reasons. First, the identifying steps are flexible because the method can be altered depending on supporting microarray platform. Another advantage is that R programming is a free software that able to analyze microarray including preprocessing, normalization and statistical detecting genes differentially expressed comparable to commercial software. Last, the proposed framework is able to identify rectal cancer genes corresponding with characteristic of colorectal cancer that can help improving the efficiency of cancer management.