CHAPTER V

RESULTS

From the two-color microarray, we analyzed the gene expression of 6 stage III rectal cancer samples and aimed to identify gene differential expression between tumor and normal tissues. Furthermore, we also considered whether the expression of these genes were up- or down-regulated. The results can be separated into three sections; the results of data before pre-processing, data pre-processing, and detecting genes differential expression according to the t-statistic.

5.1 Data before pre-processing

Each set of microarray data of rectal cancer contains an enormous amount of gene expressions, 36481 according to whole human genome. The raw data obtained from microarray experiments were in the form of images. The microarray images had to be analyzed to identify and quantify each spot and printed into intensity values. We read raw microarray data into RG values that include values of R, Rb, G, and Gb. The visualization provided already in R programming was used to check six microarray raw data. Table 5.1 shows the data of ten genes which include red and green foreground intensities further red and green background intensities. The raw intensity plot is given in Figure 5.1. In this plot, the left column shows a quick summary of red and green intensity values with a box plot and the right column shows the intensity plot of six arrays. The intensity values of six arrays varied of which the first array appeared to be more differentiated.
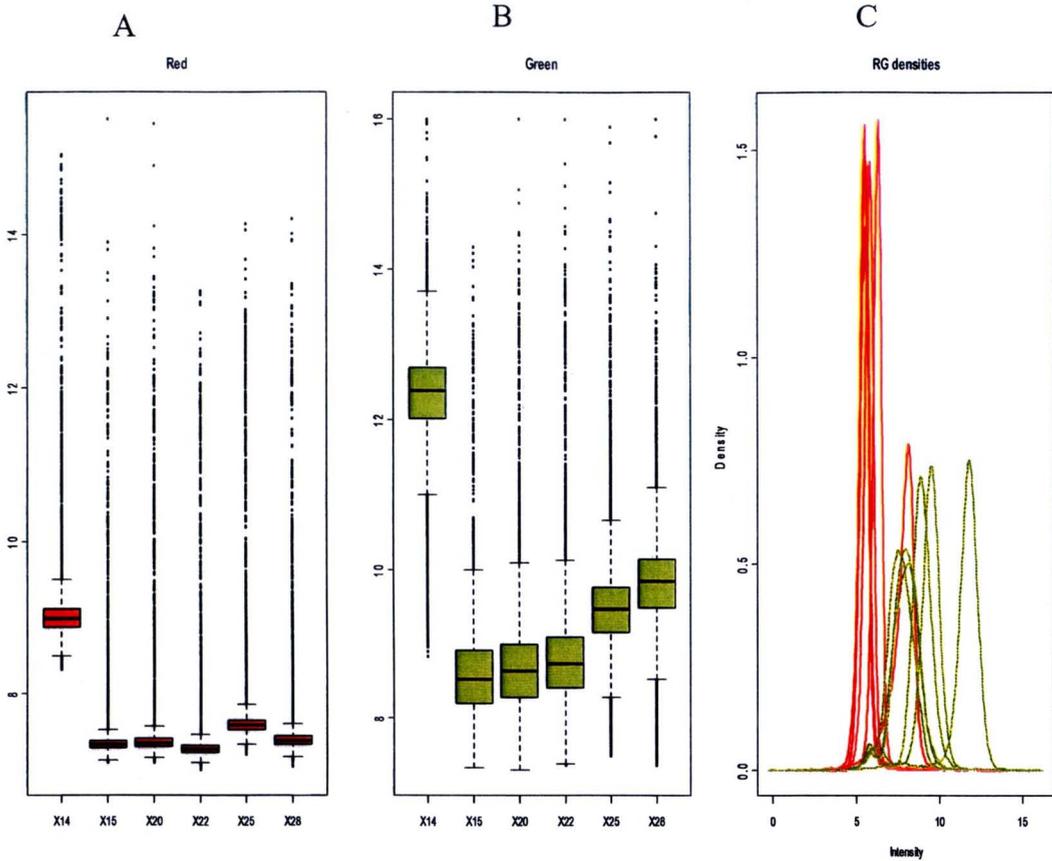
Figure 5.1 The visualization of six microarray raw data.

A) Box-plot showing the distribution and variation of intensity values form red channel of six arrays. The x-axis is showing the index of six array and y-axis showing intensity values of red channel.

B) Box-plot showing the distribution and variation of intensity values form green channel of six arrays too. The x-axis is showing the index of six array and y-axis showing intensity values of red channel.

C) Density plot showing the distribution of red and green intensity of six arrays that the distributions of red intensities are highly that than green intensities.

Table 5.1 Foreground and background intensities of red and green channels of ten genes on top of the table.

R					
Array 1	**Array 2**	**Array 3**	**Array 4**	**Array 5**	**Array 6**
24637	5496	4354	5463	9333	8069
559	158	170	163	190	162
448	180	157	164	217	166
438	156	147	164	193	166
511	171	165	183	183	183
523	161	157	150	191	167
553	157	175	167	163	168
675	163	166	159	209	182
441	167	162	159	185	166
464	148	161	144	177	175
Rb					
Array 1	**Array 2**	**Array 3**	**Array 4**	**Array 5**	**Array 6**
215	118	118	113	114	121
166	118	116	113	122	111
171	118	122	114	119	110
171	119	120	113	121	111
161	119	118	113	117	114
172	119	117	115	113	111
144	122	117	114	111	113
146	117	119	115	114	107
173	119	116	117	112	115
182	119	117	112	115	121

Table 5.1 Foreground and background intensities of red and green channels of ten genes on top of the table (continued).

G					
Array 1	**Array 2**	**Array 3**	**Array 4**	**Array 5**	**Array 6**
24509	12641	13392	14972	14730	9186
4602	206	245	202	266	205
530	179	184	183	201	182
3714	226	282	205	292	221
3028	206	212	215	324	231
3408	243	276	262	379	221
2648	208	217	259	336	210
4046	260	273	279	559	401
3036	247	278	299	635	451
2596	248	243	260	620	468
Gb					
Array 1	**Array 2**	**Array 3**	**Array 4**	**Array 5**	**Array 6**
533	124	135	133	130	128
426	119	131	138	151	124
384	123	131	129	142	119
406	124	140	131	148	127
406	119	135	130	172	143
455	123	121	136	170	142
380	124	134	131	186	129
379	132	123	133	160	138
370	119	137	133	158	135
374	123	134	140	178	146
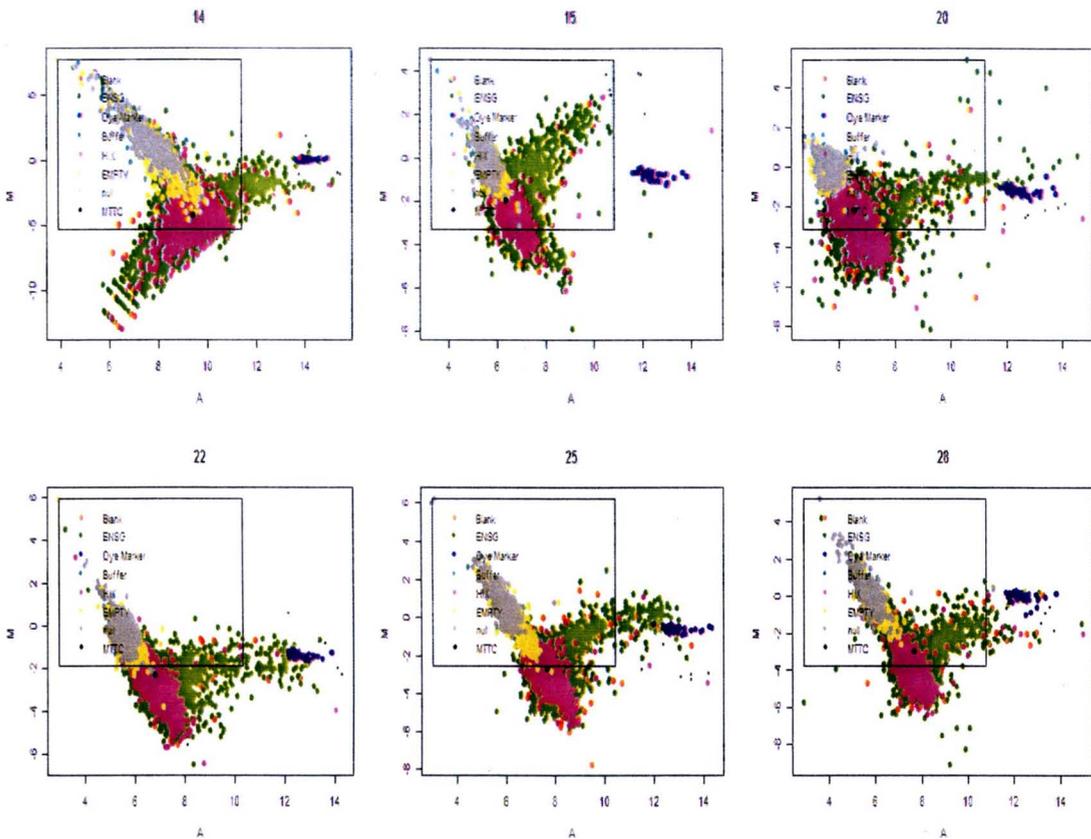
Figure 5.2 Varying MA-plot of six raw array (un-normalized data) which the number of the top MA-plot that showing the index of code microarray slides. The quality assessment of six arrays shown by MA-plots with the data un-normalized of each array is shown in Figure 5.2. The MA-plot of raw data demonstrated medium-quality data with three long comet-like patterns of non-differentially expressed probes and variety proportion values of high differentially expressed probe. Most spots were grouping with unique types such as the ENSG spot type that presented as green color grouping between negative and positive values.

5.2 Data pre-processing

Following the diagnose plot of six microarray raw data, the next process was pre-processing data. Background correction method, the normexp model, was used in the first pre-processing step. The results from the background correction are shown in the Table 5.2.

Table 5.2 Ten result values of red and green intensities after using normexp model for background correction.

R					
Array 1	**Array 2**	**Array 3**	**Array 4**	**Array 5**	**Array 6**
24271.16	5344.817	4206.683	5318.147	9163.759	7908.895
242.814	7.751788	24.68293	18.14703	13.17036	12.05466
138.7855	28.81738	6.877331	18.14703	42.75871	16.90656
131.3106	5.831472	3.249386	19.14703	16.86183	15.91591
201.5421	18.81986	17.68669	38.14703	11.48575	29.89518
202.478	9.307227	10.88812	4.066221	22.76535	16.90656
258.5391	4.831471	28.68292	21.14703	4.232037	15.91591
378.1587	12.90877	17.68669	12.14784	39.75871	35.89517
132.0436	14.84928	16.69042	10.15453	17.8278	12.05466
142.6413	2.896501	14.70934	2.582224	8.586191	14.9316

Table 5.2 Ten result values of red and green intensities after using normexp model for background correction (continued).

G					
Array 1	**Array 2**	**Array 3**	**Array 4**	**Array 5**	**Array 6**
20092.15	12399.56	13123.97	14706.54	14220.85	8342.343
1114.822	24.89596	36.66027	23.70175	53.50993	81.57775
356.9637	18.96172	23.66787	22.15221	47.42303	80.0081
818.2014	28.83392	46.31428	25.43308	57.20287	82.74524
657.3185	24.89596	27.77749	27.57863	58.29679	82.2028
728.3945	34.81869	51.9185	38.51552	67.18955	81.40069
592.5427	24.20408	28.9818	39.19699	58.02009	81.57775
926.1058	38.00759	49.66959	46.10481	118.4464	101.0019
666.1393	38.00759	45.9151	55.6275	154.8657	108.1743
584.8716	36.76997	35.23419	36.56529	137.0934	109.0391

The second step of pre-processing was normalization which was separated into two steps, within-array and between-array normalization. In the within-array normalization, loess method was used. The loess method normalized gene intensity of background correction different value in a log 2 ratio known as M-value. This was followed by between-array normalization. The Aquantile method was used to adjust the A-value base on a quantile algorithm. The results of the normalization method were reported on M and A values (Table 5.3). Figure 5.3 is quick summary plot of M and A values including the density plot of red and green channels. The box plots of M-values shows the average of all probes which was not differentially expressed.

The density plot shows the distribution of red and green channels that were not different. The results before normalization and after normalization are shown in Figure 5.4 and 5.5 respectively for comparison.

Table 5.3 M and A values of normalization method.

M					
Array 1	**Array 2**	**Array 3**	**Array 4**	**Array 5**	**Array 6**
-0.25996	-3.37235	-1.55813	-0.45728	-1.62469	-0.54569
1.018323	0.874259	1.884889	2.663635	1.783367	1.779468
1.893396	2.96478	1.577926	2.780267	2.711875	1.91514
0.105486	0.324005	-0.39417	2.613952	1.735814	1.812847
1.039302	1.907853	1.735196	3.582847	1.507503	2.148946
0.922622	0.419698	0.079207	0.342516	1.4812	1.868727
1.589087	0.498151	2.384135	2.134791	0.840463	1.851067
2.132993	0.767713	0.941154	1.080361	1.28404	2.015046
0.453342	1.021087	0.892367	0.549856	-0.36918	1.037692
0.777009	-0.79355	1.097246	0.042613	-0.71285	1.071584
A					
Array 1	**Array 2**	**Array 3**	**Array 4**	**Array 5**	**Array 6**
13.16779	13.68754	13.08791	13.2805	13.43453	13.23529
6.833251	4.77348	5.510608	5.21941	4.777519	5.021855
5.27657	5.510181	4.380448	5.179416	5.520284	5.27209
6.006018	4.673489	4.327411	5.297321	4.9948	5.25156
6.215626	5.401787	5.106397	5.761407	4.740559	5.770067
6.319556	5.138754	5.199924	4.637836	5.316169	5.284804
6.355854	4.407504	5.45556	5.61589	4.049067	5.240407
7.030456	5.432503	5.491683	5.379499	6.171991	6.105271
5.808036	5.531171	5.402763	5.383719	5.756675	5.243548
5.754499	4.340208	5.14275	4.370775	5.128505	5.423688

A   B   C
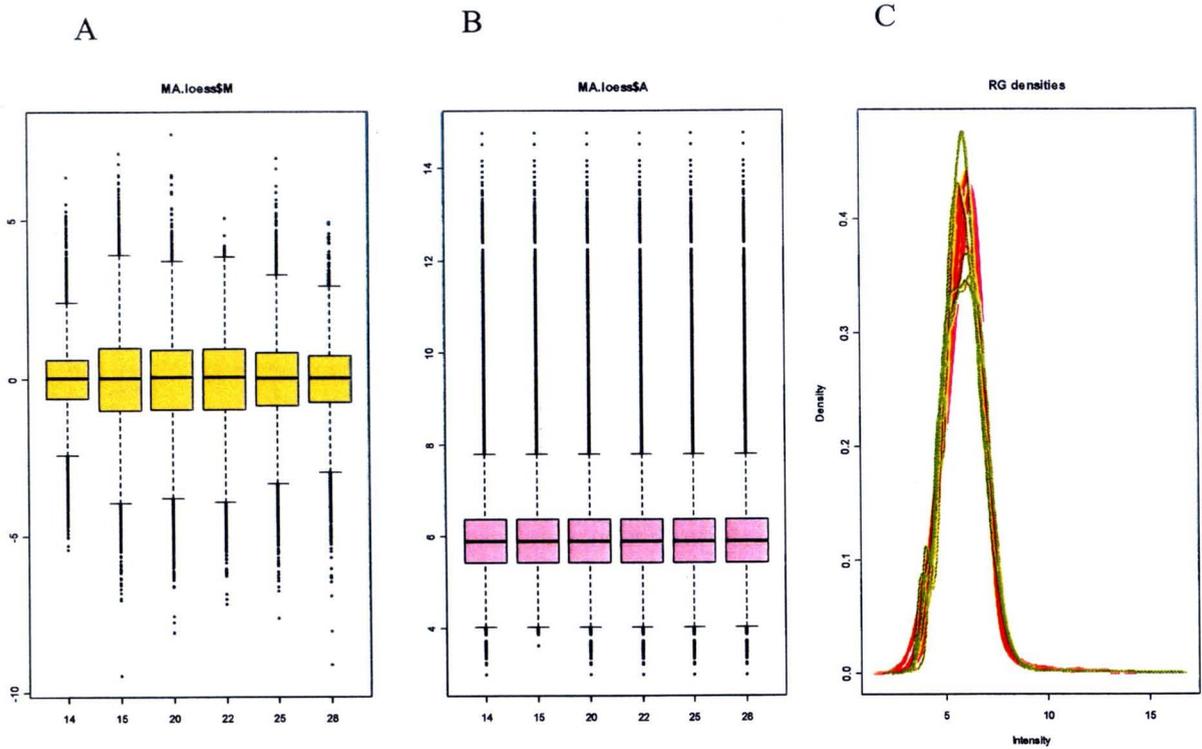
**Figure 5.3** The visualization of M-and A-values after normalization procedure.

A) Box plot of M-values from six arrays showing best normalization process and mean of M-value is zero (M=0) that associated with the assumption of normalization method. The x-axis is showing the arrangement of index of six array and y-axis showing M-value.

B) Box plot of A-values of six arrays are standard which mean values of them are six. The x-axis is showing the arrangement of index of six array and y-axis showing A-value.

C) The density plots of red and green intensities showing them are equal distribution.
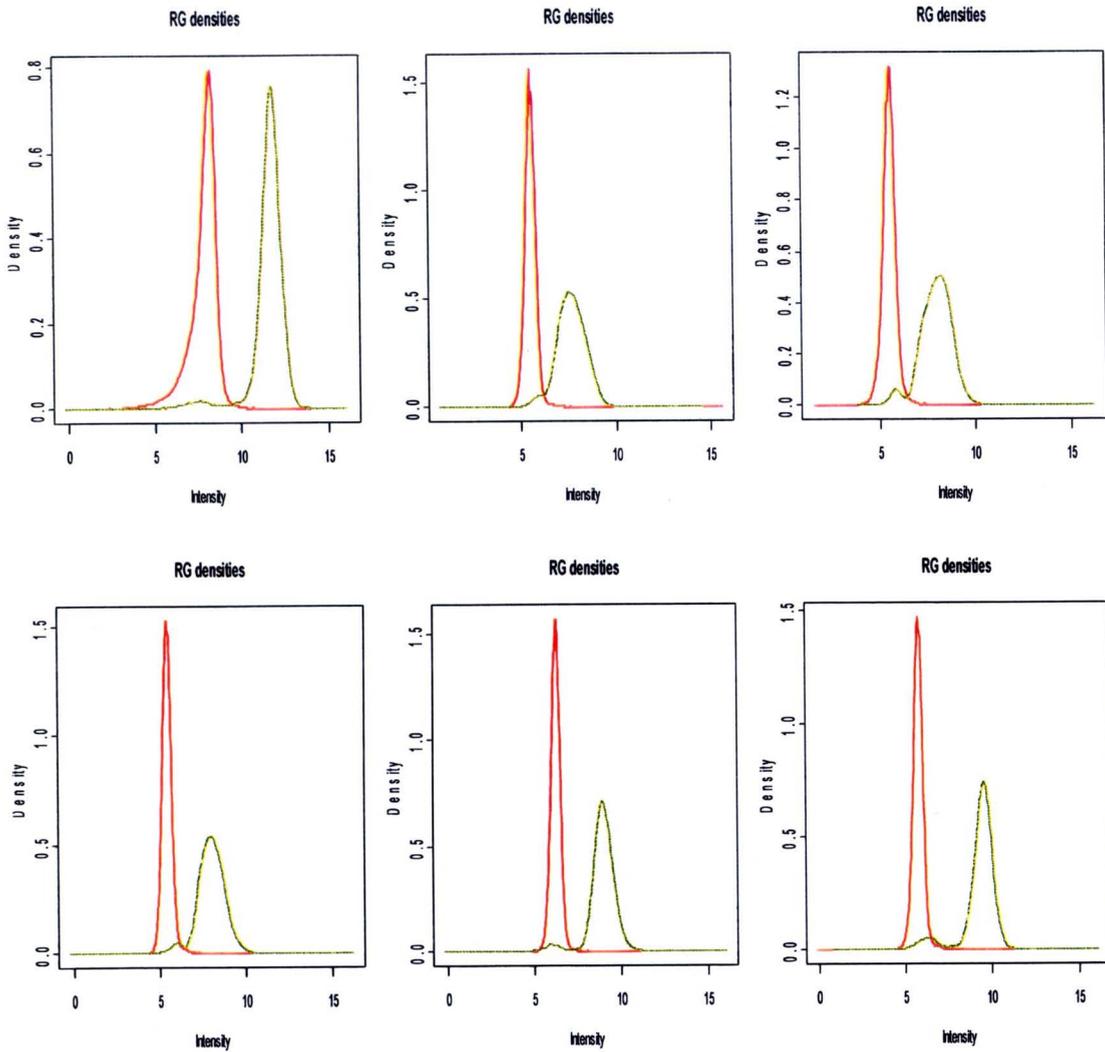
Figure 5.4 The density plot of red and green intensity for each array before normalization procedure. The distributions of red and green intensities of six arrays are difference, and subsequently most arrays red intensity highly distributed than green intensity.
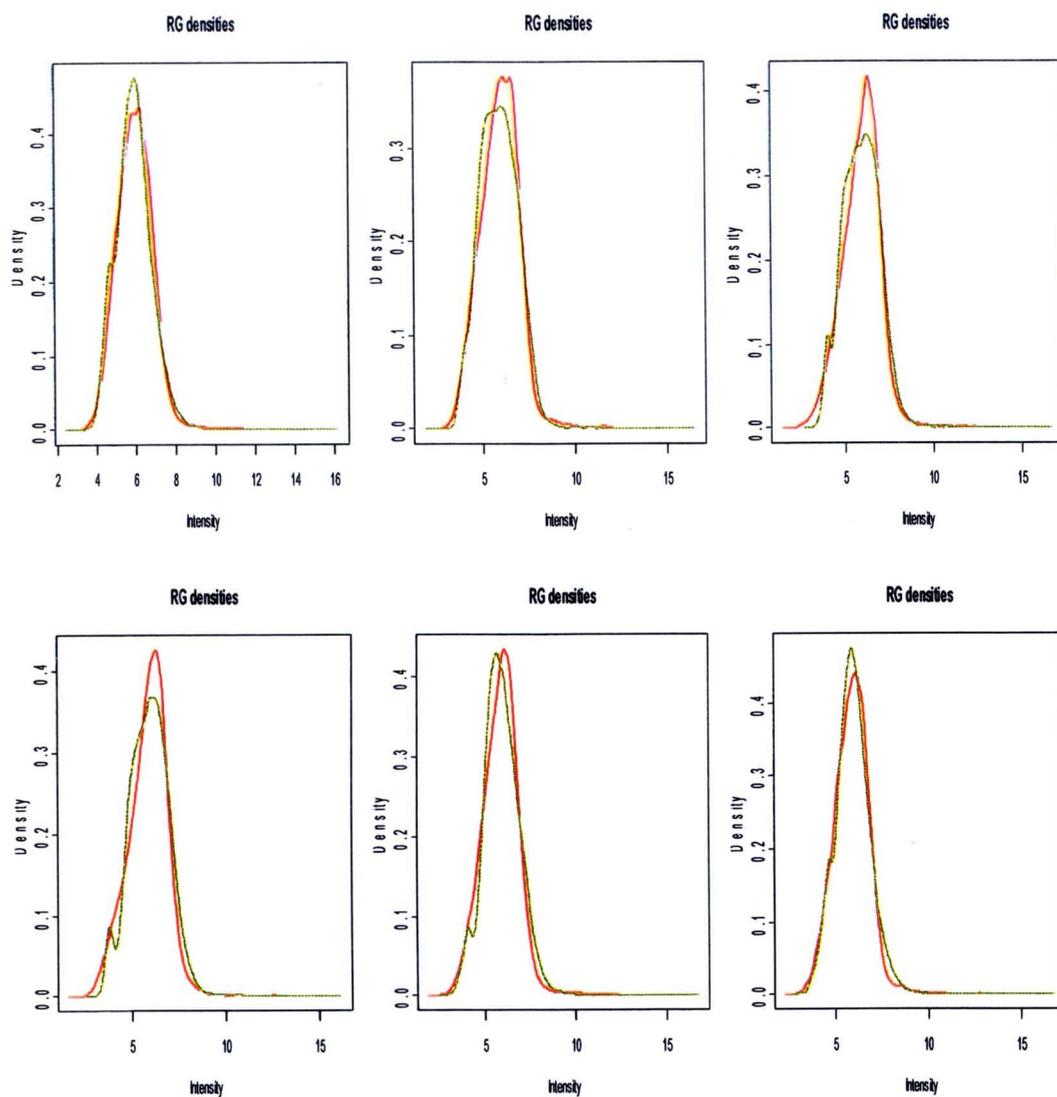
Figure 5.5 The density plot of red and green intensity for each array after normalization procedure. The distributions of red and green intensities of six arrays are not difference. Thus, six intensity plots showing the data of six arrays are normalized, that the values of red and green intensity across six arrays are standard.

5.3   Gene Differential Expression

The expression profile of six adenocacimomas of stage III rectal cancer were compared with six normal rectal samples. Differences in gene expression between normal and cancer tissues were from 6 pooled microarray data. The 304 spots were identified with significant differential gene expression as $p<0.005$ for t-test by the criteria of more than 2.0 fold change. Of the 304 spots, the blank and null status spots were filtered out resulted in 256 genes. Finally, 12 genes were identified with a clinical characteristic of colorectal cancer. Among these 12 genes, 8 genes were down-regulated and 4 genes were up-regulated. The information of 12 genes are shown in Table 5.4.

Table 5.4 Twelve genes differentially expressed at less than p-value of 0.005, comparing normal and cancer cells from the pooled data set.

Gene_id	gene_symbol	description	P.Value	logFC	Function	Status
ENSG00000164879	CA3	Carbonic anhydrase III (EC 4.2.1.1) (Carbonate dehydratase III) (CA- III). [Source:Uniprot/SWISSPROT; Acc:P07451]	0.004269	-3.79955	metabolism	Down
ENSG00000135624	CCT7	T-complex protein 1, eta subunit (TCP-1-eta) (CCT-eta) (HIV-1 Nef interacting protein). [Source:Uniprot/SWISSPROT; Acc:Q99832]	0.004617	-3.5076	protein folding/nucleotide binding	Down
ENSG00000050165	DKK3	Dickkopf related protein-3 precursor (Dkk-3) (Dickkopf-3) (hDkk-3) (UNQ258/PRO295). [Source:Uniprot/SWISSPROT; Acc:Q9UBP4]	0.002659	-3.81701	Wnt receptor signaling pathway/electron transport	Down
ENSG00000165891	E2F7	E2F transcription factor 7 [Source:RefSeq_peptide;Acc:N P_976328]	0.00306	-3.68497	transcription factor complex/regulation of transcription	Down
ENSG00000164120	HPGD	15-hydroxyprostaglandin dehydrogenase [NAD+] (EC 1.1.1.141) (PGDH). [Source:Uniprot/SWISSPROT; Acc:P15428]	0.003684	4.478824	Cellular metabolism	Up

Table 5.4 Twelve genes differentially expressed at less than p-value of 0.005, comparing normal and cancer cells from the pooled data set (continued).

Gene_id	gene_symbol	description	P.Value	logFC	Function	Status
ENSG00000172201	ID4	DNA-binding protein inhibitor ID-4. [Source:Uniprot/SWISSPROT; Acc:P47928]	0.00154	4.220407	transcription factor	Up
ENSG00000108244	KRT23	Keratin, type I cytoskeletal 23 (Cytokeratin 23) (K23) (CK 23). [Source:Uniprot/SWISSPROT; Acc:Q9C075]	0.00252	-4.13044	cytoskeleton	Down
ENSG00000133424	LARGE	Glycosyltransferase-like protein LARGE (EC 2.4.-.-) (Acetylglucosaminyltransferase-like protein). [Source:Uniprot/SWISSPROT; Acc:O95461]	0.004563	-3.76618	acetylglucosaminyltransf erase activity/glycosphingolipi d biosynthesis	Down
ENSG00000107317	PTGDS	Prostaglandin-H2 D-isomerase precursor (EC 5.3.99.2) (Lipocalin-type prostaglandin D synthase) (Glutathione-independent PGD synthetase) (Prostaglandin-D synthase) (PGD2 synthase) (PGDS) (Beta-trace protein) (Cerebrin 28). [Source:Uniprot/SWISSPROT	0.002217	-4.4632	regulation of circadian sleep/wake cycle/transporter activity	Down

Table 5.4 Twelve genes differentially expressed at less than p-value of 0.005, comparing normal and cancer cells from the pooled data set (continued).

Gene_id	gene_symbol	description	P.Value	logFC	Function	Status
ENSG00000147403	RPL10	60S ribosomal protein L10 (QM protein) (Tumor suppressor QM) (Laminin receptor homolog). [Source:Uniprot/SWISSPROT; Acc:P27635]	0.001079	4.605627	protein biosynthesis	Up
ENSG00000140988	RPS2	40S ribosomal protein S2 (S4) (LLRep3 protein). [Source:Uniprot/SWISSPROT; Acc:P15880]	0.003967	4.063239	protein biosynthesis/RNA binding	Up
ENSG00000025772	TOMM34	Mitochondrial import receptor subunit TOM34 (Translocase of outer membrane 34 kDa subunit) (hTom34). [Source:Uniprot/SWISSPROT; Acc:Q15785]	0.003428	-3.49037	protein folding/protein-mitochondrial targeting	Down