Pertanika Journal Science & Technology 25 (S): 59 - 68 (2017)

Performance Comparison of Image Normalization Method for DNA Microarray Data

Omar Salem Baans[a], Asral Bahari Jambek[a], Uda Hashim[b], Nor Azah Yusof[c].

[a]*School of Microelectronic Engineering, University Malaysia Perlis, 02600 Arau, Perlis , Malaysia*
[b]*Insituite of Nano Electronic Engineering, University Malaysia Perlis, 02600 Arau, Perlis, Malaysia*
[c]*Department of Chemistry, Faculty of Science, University Putra Malaysia (UPM), 43400 Serdang, Selangor, Malaysia*

ABSTRACT

Normalization is a process of removing systematic variation that affects measured gene expression levels in microarray experiment. The purpose is to get a more accurate DNA microarray result by deleting the systematic errors that may have occurred during the making of DNA microarray slid. In this paper, four normalization methods of Global, Lowess, Quantile and Print-tip are discussed, tested and their final results are compared in the form of Matrixes and graphs. An ideal and real microarray slides have been used for this project. It was found that the Print-tip normalization method showed the closest results to the real result for an ideal microarray slide and it has a straight median line final graph. The Print-tip normalization method uses more than one normalization factors which will be divided among intervals that are dependent on the values of the addition of red and green logarithm.

Keywords: DNA; Microarray; Normalization; Global; Lowess; Quantile; Print-tip; Background correction; M-A plot.

I. INTRODUCTION

Gene expression measurements provide clues about the regulatory mechanism, biochemical pathways and broader cellular function. By gene expression we can understand the transformation process of gene's information into proteins. The formal transformational pathway of protein begins from DNA (deoxyribonucleic acid) which is copied to the mRNA (messenger ribonucleic acid) and, finally this molecule passes from nucleus to cytoplasm carrying the information to build up proteins (Belean et al. 2011).

There are many microarray analysis software packages that are available on the market whether commercial or freeware. Basically each software program can be separated into three main tasks: The first is the gridding or addressing, which is the process of specifying coordinate to every spot on the slide. Secondly, the segmentation which decides the classification of each pixel either as foreground which corresponds to be an interest spot or as background which acts as an error or noise. The third and the last task is the Intensity Extraction which is the step to calculate green and red for foreground fluorescence intensity for each spot on the array (Borda M *et al.* 2011), (Rao Y *et al.* 2008).

Subsequently, there are many processes to inspect the results and also to correct the errors that have occurred. The first is the background correction method; ignoring the effect of intensity of the background. This can be achieved by subtracting the value of the background intensity from the value of foreground intensity or any other suitable method to neglect the effect of background intensity. Another process to increase the accuracy is the normalization method which we are going to discuss in this paper (Yang Y *et al.* 2001).

Normalization: a process of removing systematic variations that affect measured gene expression levels in microarray experiments. The purpose of normalization is to adjust for effects which arise from variations in the microarray technology rather than from biological differences between the RNA samples or between the printed probes. Imbalances between the red and green dyes may arise from differences between the labeling efficiencies or scanning properties of the two flours complications perhaps by the use of different scanner settings (Geeleher P *et al.* 2009). The aim of this paper is to review various methods that discuss DNA microarray normalization and make comparison among them.

Section II several normalization algorithms are elaborated, while section IV discusses the comparison of these varies methods. Section V and VI represent methodology and results of each method and section VII conclude this paper.

II. LITERATURE RIVIEW

Normalization of DNA microarray has been discussed in many studies. Before we review some of them, we will explain the two types of graphs than can show normalization quality. First, (log M vs. log R) as shown in Figure 1(a). Second, M-A plot is 45° rotation of standard scatter plot as shown in Figure 1(b). Write R and G for the background-corrected red and green intensities for each spot. Normalization is usually applied to the log-ratios of expression, which will be written (M = log R – log G). The log-intensity of each spot will be written (A = (log R + log G)/2), a measure of the overall brightness of the spot. (The letter M is a mnemonic for minus while A is a mnemonic for addition) (Dudoit S *et al.* 2002).

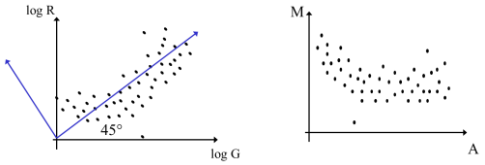


Fig. 1: (a) Log R vs. Log G; (b) M-A Plot.

This section will discuss and elaborate these methods of DNA microarray normalization in order to choose the most suitable one and develop it for further microarray analysis. The first method is Global normalization: The underlying assumption of this approach is that the total of mRNA labeled with either R value (sum of red intensities) or G value (sum of green intensities is equal. While the intensity for any one spot may be higher in one channel than the other, when averaged over thousands of spots in the array, these fluctuations should average out. Consequently, in this method, it takes the value of c out of log (R/G). The c value is equal to the main assumption that equal to log of the total R over total G which can be expressed by the variable K (Yang, Y *et al.* 2002). The intensity-dependent lowess normalization runs a line through the middle of the MA plot, shifting the M value of the pair (A,M) by c=c(A), as shown in Equation 3. One estimate of c(A) is made using the loess function: LOcally WEighted Scatterplot Smoothing (Berger J *et al.* 2004) (Bilban M J *et al.* 2002).

$$log_2 R/G \rightarrow log_2 R/G - c = log_2 R/(kg) \quad \ldots\ldots\ldots\ldots (1)$$

$$K = \sum R/G \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

$$log_2 R/G \rightarrow log_2 R/G - c(A) = log_2 R/(k(A)G) \quad \ldots\ldots\ldots (3)$$

In the Print-tip normalization, each M-value (Log R – Log G) is normalized by subtracting from it the corresponding value of the tip group loess curve that is dependent on A value ([Log R + Log G]/2) while its value should be fixed. The normalized log-ratios (N) are the residuals from the tip group loess regressions. A simpler form of Print-tip is shown in Equation 4 where loess (A) is the global loess curve plotted in Figure 3(a). Refer to Figure 3(b) for the final figure of the Print-tip normalization (Smyth G *et al.* 2003). Lastly is the Quantile normalization method which is also one of the most favorable approaches used especially in normalization between arrays. First, rearrange the genes in each column as in second table in Figure 4. Then, take the mean in each raw and replace the whole raw by the mean value as shown in the third table in Figure 4. Finally, reorder each gene in its original place with its new value.

$$N = M - loess(A) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

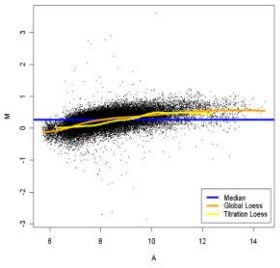


Fig. 2: Global normalization;

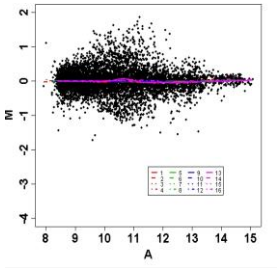


Fig. 3: Print-tip normalization;

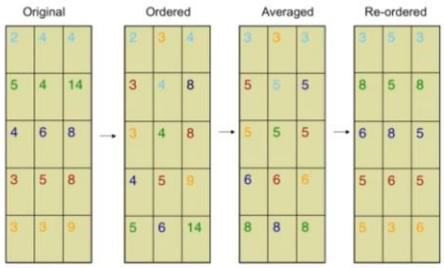


Fig. 4: Quantile normalization.

III. COMPARISON OF DIFFERENT NORMALIZATION APPROACHES

In this section, the existing system algorithm as discussed in section III will be analyzed and discussed to find out the similarities and variations among the different normalization methods. Table 1 summarized the comparison of these algorithms.

From table 1, it can be seen that, all the methods are using mainly the value of M which equal to log of red intensity minus log of green intensity. However, three methods have different value to subtract from M. To illustrate, Global normalization use the log of addition of each of red and green intensity while the other two methods are using median and global median.

In term of the final shape of the normalization on M-A graph, there are similarities between Lowess and Print-tip methods because both have a straight median line in the value of (M = 0) due to their similarities on subtracting the mean or median from M. However, in Global normalization, there is a curve around the value of (M= 0) due to the subtraction of the total R and G. moreover, Quantile normalization method does not use M-A plot, consequently its final graphs does not always take a straight line of the mean on the (M=0). According to this review, we suggest Print-tip normalization method to be used because when comparing to the global normalization its final figure is simpler and easier to read and can also easily be compared to various plots. Straight line on (M=0) is easier to read than the Global and lowess normalization curve.

No.	[1]	[2]	[3]	[4]
Method	Global	Lowess	Print Tip	Quantile
Function	Log (R/ KG)	Log (R/ G) – c(A)	N= M- loess (A)	Mean of rows after reorder
Variable	$K = \sum R/G$	LOWESS function	Global Loess	NA
Shape on M-A graph	Curve	Straight line on (M=0)	Straight line on (M=0) but has some variation	It does not meet M-A plot.

Table 1 Comparison between different system algorithms

IV. METHODOLOGY

Using Matlab, we developed a code that can extract the intensity for 100 spots. Using 100 spots instead of the whole microarray slide make the process easier and simpler especially to compare the many algorithms used. In order to examine the suitable method which would be more accurate and suitable for this project, an ideal microarray image spots in Figure 5(a), and a real microarray slide in Figure 5(b) were used. Matlab usually reads the image intensity as matrix by pixel, for example our image after cropping is 220*227 pixels while it has only 100 spots. Thus each spot has around 20 pixel diameters. Next, it calculate the foreground and background then subtract the background value from foreground, and using threshold equal to zero will not allow negative values to appear. In the ideal image the value of background is fixed (Rb = Gb =3) while foreground value is a variant from 0 to 225 as shown in Matrix 1. Then, according to the normalization method, the formula codes were applied.

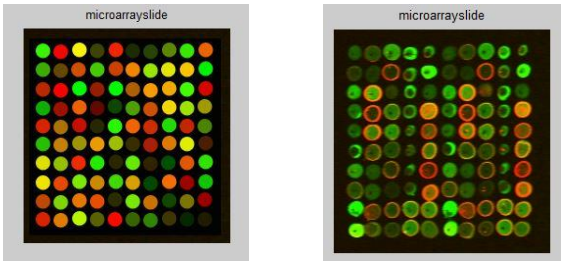


Fig. 5: (a) Ideal microarray slide with 100 spot; (b) Real microarray slide with 100 spot;

RedIntensity =

51	243	250	57	242	25	36	91	51	233
57	99	207	67	201	239	148	227	246	1
184	251	1	149	1	172	247	238	65	210
10	155	234	95	14	73	203	239	154	157
214	170	194	36	36	238	201	28	197	73
35	193	155	233	154	54	169	218	219	71
219	147	238	32	39	96	46	1	222	44
236	228	121	203	63	212	1	244	156	15
190	124	230	228	42	117	201	14	121	150
214	220	180	86	249	51	46	60	5	28

GreenIntensity =

244	3	241	66	32	40	66	132	231	96
170	69	76	199	61	129	226	225	192	237
36	1	251	27	251	92	152	166	248	3
172	14	94	6	66	157	77	212	224	150
31	112	12	38	245	96	43	216	33	128
190	91	144	151	188	42	27	118	235	27
220	191	40	230	36	232	33	78	85	224
224	70	230	133	213	157	69	106	1	193
68	210	115	74	40	230	111	68	114	1
32	121	243	94	4	97	130	50	35	25

Matrix 1: Original Intensity of the ideal spots.

For Global Normalization, loops were used to find the total of red and green intensities values for all 100 spots. Then taking the logarithm of the total value and subtracts it from the value of M according to Equation number 1 and 2. Similarly, in the Lowess method, mean of m values was calculated then subtracted, to be on the center (M=0) according to equation number 3. However, Quantile normalization is much different than the previous two methods, because it does not require calculation of A and M values. But it requires sorting the matrix in each column. Then taking the average in each raw and finally put each new value in its original

location as shown in Figure 5. Finally, Print-tip normalization, A values (addition of logarithm) has been divided into four groups (<5, <6, <7 and else) and according to each group, mean value of M was taken and defined into variable call PT. After that, the PT value was subtracted from M according to its group. Next section will discuss the results of the various methods tested.

V. RESULT AND DISSCUSSION

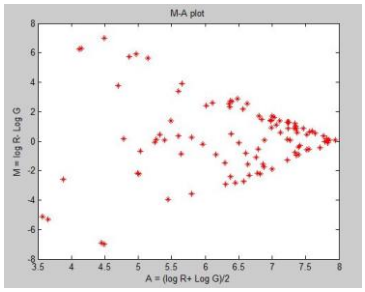
First of all, there is a different in the last result for all the four methods in terms of last intensities values and M-A displaying graphs. Global normalization and Lowess share a similarity especially when we compare the difference between the green and the red intensity for the same spots. Similarly, Print-tip normalization which has a similar graph but there is a different according to the interval groups. However, the results for quantile normalization are fluctuating and the different is larger than all of the other normalization methods. Normalization results for the ideal and normal microarray slide are shown in matrix 1 and 2, and M-A graphs in Figure 6 and 7 respectively. As we saw in Matrix 1 above, there are red and green intensities for 100 spots as well as in Matrix 2 bellow. Thus, we have 4 matrixes with the size of (10*10). The first and second for the red and green intensities of ideal image in Matrix 2 while the third and fourth for the red and green intensities for the slide image. Figure 6 and 7 depict M-A plots for ideal and slide image before any method of normalization was performed. Thus, the illustrations will help us compare them with the next results of various normalization methods.

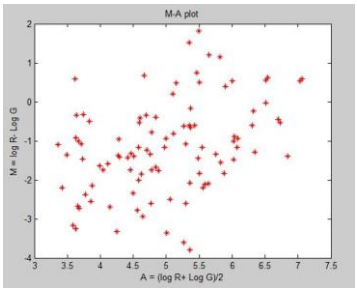
RedIntensity =

5	15	22	11	11	4	16	9	12	14
11	6	36	19	30	8	4	41	20	22
10	87	13	12	5	8	88	31	16	3
5	56	23	15	156	8	75	19	11	163
4	63	21	8	67	3	89	33	11	76
8	32	11	6	47	13	26	20	11	73
13	6	20	7	68	15	5	24	8	84
15	9	6	10	109	32	14	8	9	115
10	53	25	32	44	10	83	38	48	45
70	34	13	23	30	51	30	22	38	26

GreenIntensity =

30	50	99	39	16	31	42	102	46	9
31	30	31	28	110	27	22	29	27	43
66	126	26	34	30	68	120	19	39	37
59	33	85	35	108	61	32	83	30	108
32	96	100	17	51	35	91	85	16	52
29	51	47	14	88	36	56	46	14	86
78	44	35	21	23	96	34	32	16	23
53	50	17	13	74	49	49	18	29	74
152	37	45	96	100	133	37	43	93	91
185	52	34	102	69	126	47	28	108	73





Matrix 2: Red and Green Intensity before norm of the slide image; Figure 6: M-A plot before normalization of the ideal image; Figure 7: M-A plot before normalization of the slide image.

Matrixes 3 and Matrix 4 show the results of global normalization for ideal and real DNA microarray slide. Firstly they show k and c values, c is the logarithm of the total of red intensities over the total of green intensities (k) which is equal to 0.0274 in ideal image and – 0.2358, and that explains to us why the normalization is important and how the variety of c increased for the real microarray slide image. Thus, the difference between the last and original results in the real microarray slide is larger.

k =

1.0277

c =

0.0274

RedGlobNorm =

52	242	249	58	241	26	37	92	52	232
58	100	207	68	201	238	148	226	245	2
184	250	2	149	2	172	246	237	66	210
11	155	233	96	15	74	203	238	154	157
213	170	194	37	37	237	201	29	197	74
36	193	155	232	154	55	169	217	218	72
218	147	237	33	40	97	47	2	221	45
235	227	121	203	64	211	2	243	156	16
190	124	229	227	43	117	201	15	121	150
213	219	180	87	248	52	47	61	6	29

GrenGlobNorm =

248	5	245	68	34	42	68	135	235	98
173	71	78	202	63	132	230	229	195	241
38	3	255	29	255	94	155	169	252	5
175	16	96	8	68	160	79	216	228	153
33	115	14	40	249	98	45	220	35	131
193	93	147	154	191	44	29	121	239	29
224	194	42	234	38	236	35	80	87	228
228	72	234	136	217	160	71	109	3	196
70	214	118	76	42	234	114	70	117	3
34	124	247	96	6	99	133	52	37	27

k =

0.7900

c =

-0.2358

RedGlobNorm =

7	18	25	14	14	6	19	11	15	17
14	8	41	22	34	10	6	46	23	25
12	96	16	15	7	10	97	35	19	5
7	62	27	18	171	10	83	22	14	178
6	70	24	10	74	5	98	37	14	84
10	36	14	8	53	16	30	23	14	81
16	8	23	9	75	18	7	28	10	93
18	11	8	12	120	36	17	10	11	126
12	59	29	36	49	12	92	43	54	50
78	38	16	27	34	57	34	25	43	30

GrenGlobNorm =

29	47	93	37	16	30	40	95	44	10
30	29	30	27	103	26	22	28	26	41
62	118	25	33	29	64	112	19	37	36
56	32	80	34	101	58	31	78	29	101
31	90	94	17	48	34	85	80	16	49
28	48	45	14	83	35	53	44	14	81
73	42	34	21	23	90	33	31	16	23
50	47	17	13	70	47	47	18	28	70
141	36	43	90	94	124	36	41	87	85
172	49	33	95	65	118	45	27	101	69

Matrix 3: Red and Green Intensity for global norm of the ideal image; Matrix 4: Red and Green Intensity for global norm of real slide image;

Lowess normalization results for ideal and real DNA microarray slide are shown in Matrix 5 and 6. First it shows (m) values, m is the mean of M values for 100 spots which equal to the difference between logarithms of red and green intensities for each spot separately. (m) is equal to 0.1756 in ideal image and –1.1662 and also explains to us why the normalization process is important and how does the variety of c increase for the real

microarray slide image. Also it is greater than c values (for global normalization). Thus, the different between the last and original results in real microarray slide is larger and this is larger than the different in global normalization.

m =

0.1756

RedlowsNorm =

49	230	237	55	229	25	35	87	49	221
55	95	196	64	191	226	141	215	233	2
175	238	2	142	2	163	234	225	63	199
11	147	222	91	15	70	192	226	146	149
203	161	184	35	35	225	191	28	187	70
34	183	147	221	146	52	160	207	208	68
208	140	225	32	38	92	45	2	210	43
224	216	115	192	61	201	2	231	148	16
180	118	218	216	41	112	191	15	115	143
203	208	171	82	236	49	45	58	6	28

GrenLowsNorm =

261	5	258	72	36	44	72	142	247	104
182	75	82	213	66	139	242	241	206	253
40	3	268	30	268	99	163	178	265	5
184	16	101	8	72	168	83	227	240	161
35	121	14	42	262	104	47	231	37	138
203	98	155	162	201	46	30	127	251	30
235	205	44	246	40	248	37	84	92	240
240	76	246	143	228	168	75	114	3	207
74	225	124	80	44	246	120	74	123	3
36	130	260	101	6	105	140	55	39	28

Matrix 5 R & G Intensity for Lowess norm of the ideal image;

m =

-1.1662

RedlowsNorm =

9	24	35	18	18	8	26	15	20	23
18	11	56	30	47	14	8	63	32	35
17	132	21	20	9	14	134	48	26	6
9	86	36	24	236	14	114	30	18	246
8	96	33	14	102	6	135	51	18	116
14	50	18	11	72	21	41	32	18	111
21	11	32	12	104	24	9	38	14	128
24	15	11	17	165	50	23	14	15	174
17	81	39	50	68	17	126	59	74	69
107	53	21	36	47	78	47	35	59	41

GrenLowsNorm =

21	35	67	27	12	22	29	69	32	7
22	21	22	20	75	19	16	21	19	30
45	85	19	24	21	47	81	14	27	26
41	23	58	25	73	42	23	57	21	73
23	65	68	13	35	25	62	58	12	36
21	35	33	11	60	25	39	32	11	59
53	31	25	15	17	65	24	23	12	17
37	35	13	10	51	34	34	13	21	51
103	26	31	65	68	90	26	30	63	62
125	36	24	69	47	85	33	20	73	50

Matrix 6: R & G Intensity for Lowess norm of real slide image;

Quantile normalization results for ideal and real DNA microarray slide are shown is Matrix 7 and 8. Quantile normalization method is much different from the previous methods because it does not require fixed values of (c) or (m) like in global and Lowess normalizations. It takes an average of the columns after sorting the matrix in each raw as explained before in section 2. Thus, we can see in Matrix 8 that (67, 85, 124, 18 and so on) are repeated in each column of matrix QRN, and also the anther values for QGN are similar in Matrix 8. There are 10 fixed numbers repeated in each column of each matrix.

QRN =

67	229	243	67	229	18	50	124	50	243
85	18	149	85	211	243	124	192	243	18
124	243	18	192	18	192	243	211	67	229
18	85	211	149	50	85	229	229	124	211
192	124	124	50	67	229	192	67	192	149
50	149	67	243	192	67	149	149	211	124
229	67	229	18	85	124	67	18	229	85
243	211	50	211	149	211	18	243	149	50
149	50	192	229	124	149	211	50	85	192
211	192	85	124	243	50	85	85	18	67

QGN =

236	32	187	72	32	18	72	123	187	99
99	72	48	216	99	123	236	236	123	236
48	18	236	32	236	48	216	167	236	32
123	48	72	18	123	167	123	187	167	167
18	167	18	48	216	72	48	216	32	123
167	123	123	187	167	32	18	99	216	72
187	216	32	236	48	236	32	48	72	216
216	99	167	167	187	187	99	72	18	187
72	236	99	99	72	216	167	32	99	18
32	187	216	123	18	99	187	18	48	48

Matrix 7: R & G Intensity for Quant. Norm. of the ideal image;

QRN =

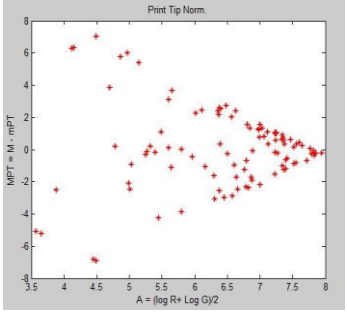
9	16	38	23	9	9	16	9	28	9
38	6	78	42	14	14	6	78	42	14
23	78	14	28	6	16	55	38	38	6
14	42	42	38	78	23	38	14	14	78
6	55	28	14	38	6	78	42	16	38
16	23	9	6	28	38	23	16	23	28
42	9	23	9	42	42	9	28	6	42
55	14	6	16	55	55	14	6	9	55
28	38	55	78	23	28	42	55	78	23
78	28	16	55	16	78	28	23	55	16

QGN =

25	44	104	73	20	25	44	120	73	20
30	20	30	44	120	20	20	36	36	36
57	120	25	52	30	57	120	25	57	30
52	25	73	57	104	52	25	73	52	120
36	104	120	30	36	30	104	104	25	44
20	57	57	25	57	36	73	57	20	73
73	36	44	36	25	73	30	44	30	25
44	52	20	20	52	44	57	20	44	57
104	30	52	104	73	120	36	52	104	104
120	73	36	120	44	104	52	30	120	52

Matrix 8: R &G Intensity for Quant. norm of the real slide image;

Finally, Print-tip normalization gave the results for the red and green intensities for the ideal microarray image in Matrix 9 and real microarray slide in Matrix 10. Also M-A graphs for the results are displayed in Figure 8 and 9 respectively. PT values in Matrix 9 and 10 are represented by the normalization values among the four intervals for each image. For example, in Matrix 9, PT equals -0.0664, 0.2457, 0.1445 and 0.2633. These values were subtracted from M (the different between logarithms of red and green intensities for each spot) according to the values of A for the same spot. These interval are (<5, <6, <7 and else), so each interval has its own normalization values; and that is why, at times, we can see the obvious different between the normalized and un-normalized values in some intervals according to the values of PT. Besides that, Figure 9 represents the M-A plot for Print-tip normalization of real image slide which show more different from its original slide except by the values of PT especially in the first interval when PT= -1.5263 among the interval (A less than 5).



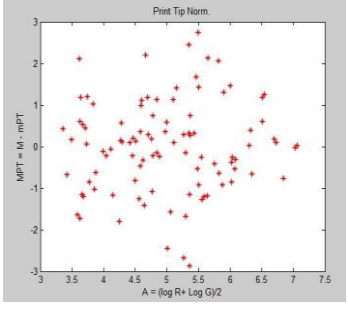


Figure 8: M-A plot for Print-tip norm of the Ideal image; Figure 9: M-A plot for Print-tip norm of the slide image

```
PT =

   -0.0664    0.2487    0.1445    0.2633


RedPTNorm =

    50   250   230    54   232    24    34    88    50   214
    56    96   198    65   193   220   137   209   226     3
   176   258     3   143     3   165   227   219    61   216
    11   144   215    99    16    71   195   220   142   145
   205   157   179    34    36   219   193    28   189    71
    35   178   143   214   142    51   162   200   201    67
   201   136   228    32    37    89    44     3   204    43
   217   218   112   187    61   195     3   224   161    15
   182   115   211   210    40   108   185    14   117   155
   205   202   166    83   230    50    45    56     7    30


GrenPTNorm =

   258     4   266    74    35    45    74   140   244   107
   180    74    81   211    66   143   249   248   212   233
    39     2   247    30   247    98   168   183   273     4
   189    17   105     7    66   167    83   234   247   166
    34   124    15    43   259   107    47   229    36   136
   201   101   159   167   208    47    30   131   259    31
   243   211    44   243    41   256    38    78    95   237
   247    75   254   147   225   174    69   118     2   212
    73   232   128    83    45   254   123    76   121     2
    35   134   268   100     6   104   138    56    36    26
```

```
PT =

   -1.5263   -0.9235   -0.6295    0.5579


RedPTNorm =

    11    28    32    21    21     9    29    14    23    26
    21    12    51    34    43    16     9    58    36    40
    19   110    24    23    11    16   111    55    29     7
    11    79    34    28   130    16   105    28    21   136
     9    80    31    16    94     7   112    47    21   107
    16    46    21    12    60    24    38    36    21    93
    20    12    36    14    96    23    11    43    16   118
    28    17    12    19   137    46    26    16    17   145
    16    75    36    46    56    16   116    54    61    58
    89    49    24    34    43    65    43    40    49    38


GrenPTNorm =

    19    31    73    24    11    19    26    75    28     6
    19    19    24    18    81    17    14    22    17    26
    40   103    16    21    19    41    98    12    24    23
    36    25    63    22   133    37    24    61    19   133
    20    78    74    11    38    22    74    63    11    39
    18    38    29     9    72    22    42    28     9    70
    58    27    22    13    18    71    21    20    11    18
    32    31    11     9    61    37    30    12    18    61
   112    28    34    71    82    98    28    32    76    74
   150    39    21    75    51   103    35    18    88    54
```

Matrix 9: Red and Green Intensity for PT norm of the ideal image; Matrix 10: Red and Green Intensity for PT norm of real slide image;

From the Matrixes and graphs discussed above, it can be noticed that the global and Lowes are almost similar and Print-tip is an advanced version of two and they gave a close results to the correct one in Matrix 1 and 2. However, Quantile differed greatly than the correct one and its graphs fluctuate away from the goal. Furthermore, the graphs of real image Print-tip normalization shows the expected result for real slide image in Figure 9 due to the clustering around the straight line when (M = 0). This findings support the finding of Smyth as he mentioned that the "print-tip loess normalization provides a well-tested general purpose normalization method which gives good results on a wide variety of arrays". It is best combined with diagnostic plots of the data. When the diagnostic plots show that biases still remain in the data after normalization, further normalization steps such as quantile normalization between the arrays may be undertaken (Smyth G *et al.* 2003).

VI. CONCLUSION

In this paper, normalization is defined as a process to delete systematic error which is why it is important and necessary. Since there are many normalization methods that exist, four most commonly used normalization algorithms such as Global, Lowess, Quantile and Print-tip have been tested and compared to find the most suitable approach in a general normalization process. For that purpose, a Matlab code was built for each method for two slides; the ideal and real microarray slides. The results were shown in two forms, Matrix of red and green intensities and M-A graph. The results show that Global, Lowess and Print-tip have a more accurate result once compared with an ideal image result while Print-tip has the advantages than the other two especially in term of final graph shape.

ACKNOWLEDGMENT

This research was funded by ScienceFund, Ministry of Science, Technology and Innovation (MOSTI), Malaysia (2015).

REFERENCES

Belean, B., Borda, M., LeGal, B., & Malutan, R. (2011, August). FPGA technology and parallel computing towards automatic microarray image processing. *In Telecommunications and Signal Processing (TSP), 2011 34th International Conference on* (pp. 607-610). IEEE.

Berger, J. A., Hautaniemi, S., Järvinen, A. K., Edgren, H., Mitra, S. K., & Astola, J. (2004). Optimized LOWESS normalization parameter selection for DNA microarray data. *BMC bioinformatics*, 5(1), 1

Bilban, M., Buehler, L. K., Head, S., Desoye, G., & Quaranta, V. (2002). Normalizing DNA microarray data. *Current Issues in Molecular Biology*, 4, 57-64.

Borda, M., Belean, B., Terebes, R., & Malutan, R. (2011, November). FPGA based SoC for automated cDNA microarray image processing. *In E-Health and Bioengineering Conference (EHB)*, 2011 (pp. 1-4). IEEE.

Dudoit, S., Yang, Y. H., Callow, M. J., & Speed, T. P. (2002). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica sinica*, 111-139.

Geeleher, P., Morris, D., Hinde, J. P., & Golden, A. (2009). BioconductorBuntu: a Linux distribution that implements a web-based DNA microarray analysis server. *Bioinformatics*, *25*(11), 1438-1439.

Rao, Y., Lee, Y., Jarjoura, D., Ruppert, A. S., Liu, C. G., Hsu, J. C., & Hagan, J. P. (2008). A comparison of normalization techniques for microRNA microarray data. *Statistical applications in genetics and molecular biology*, 7(1).

Smyth, G. K., & Speed, T. (2003). Normalization of cDNA microarray data. *Methods*, *31*(4), 265-273.

Yang, Y. H., Buckley, M. J., & Speed, T. P. (2001). Analysis of cDNA microarray images. *Briefings in bioinformatics*, *2*(4), 341-349.

Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J., & Speed, T. P. (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic acids research*, *30*(4), e15-e15.