JERE, *Volume 10, No.1 (27-36),* Special Issue on Bioprocess & Biosystem Engineering and their Downstream Processing

Analysis of Background Correction Methods for DNA Microarray Imaging

Omar Salem Baans[1*] and Asral Bahari Jambek [2]

[1, 2] *School of Microelectronic Engineering, University Malaysia Perlis, 06200 Arau, Perlis, Malaysia.*

ABSTRACT

Most microarray image scanning approaches provide an estimation of the intensity of the foreground and background for each spot. Background intensity must be corrected in order to remove the effect of the non-specific binding or spatial heterogeneity across the array, but when such corrections are applied many problems appear, such as negative intensity for the spot or high variability of low-intensity log ratios. In this paper, many alternative methods for calculating background intensity are discussed and many approaches for background correction are tested and compared. GenePix, ScanAlyze and QuantArry are the methods reviewed for background intensity extraction. Four methods for background correction were examined and the Edwards method was found to be the best and most suitable method.

Keywords: Microarray Imaging, Background Correction, DNA, Edward Method.

1. INTRODUCTION

Gene expression regulates the production of proteins, which control all cellular processes in the human biological system. The understanding of gene expression and the mechanism of protein production has many applications in terms of diagnosis, staging and finding the suitable treatments for diseases. Using the cDNA microarray, it is possible to diagnose rapidly and efficiently the level of gene expression in the sample [1] and [2].

There are many commercial and freeware microarray analysis software packages available. Each software program can be separated into three main tasks. The first is gridding or addressing, which is the process of specifying coordinates for every spot on the slide. The second is segmentation, which classifies each pixel as either foreground, corresponding to a spot of interest, or as background, corresponding to error or noise. The third and final task is intensity extraction, which is the step for each spot on the array, calculates the green and red foreground fluorescence intensity [3], [4] and [5].

The estimation of the background intensity is a very important step to be performed as a part of the background correction process. This is because each spot intensity measure includes a contribution to the fluorescence that is not due to hybridization of the mRNA sample to produce spotted DNA. Background intensity can be estimated by more than one method, for example, using the main concentration of pixels located outside the spot mask [6] and [7].

The background correction for the spot intensity can usually be performed by subtracting the background intensity from the foreground intensity of the spot, but sometimes negative values appear where the background value is greater than the foreground value for a given spot, which seems illogical. Many studies have discussed this issue and proposed various methods that can be applied to avoid the problem [8] and [9].

* Corresponding Author: *omersalim4901@gmail.com*

The main aim of this paper is to discuss and compare the background correction alternatives and the algorithms that are used to calculate the background value pair (Rb, Gb). In this paper, several microarray background correction methods are discussed. Section 2.1 presents the existing algorithm systems for calculating the background intensity values, while section 2.2 discusses and compares different methods used for the background correction. Section 3 discusses the methodologies used in this project to correct background intensity and section 4 elaborates and compares the results of different methods. Finally, section 5, presents the conclusions.

2. LITERATURE REVIEW

2.1 Most Popular Background Intensity Extraction Methods

Most microarray analyses define foreground intensity values for red and green fluorescence (Rf, Gf) as the mean or median value of all the pixel values inside the segmented spot mask as shown by the red circle at the center. However, other variations are available for the calculation of the background intensity values for red and green fluorescence (Rb, Gb) especially with respect to selecting the background region. Among the most popular approaches that utilize the median for the specially-selected region around the spot mask are GenePix, ScanAlyze and QuantArry.

The GenePix package considers the median value of all the pixel values inside the valley region as the foreground intensity value of the spot. The valley region is represented in Figure 1 as four pink rectangular areas surrounding the spot [10]. The background intensity estimation methods implemented in ScanAlyze consider all the pixels that are located outside the spot mask but within the square where the spot lies. This is represented in Figure 1 as the blue square surrounding the red circle. The median of all these pixels is used as the estimated value of the local background intensity of the spot. Finally, the QuantArry method uses the median of every pixel in the area between two green concentric circles in Figure 1 as the background intensity value for that spot [11]. While the QuantArry method uses the median of every pixel in the area between the two green concentric circles in Figure 1 as the background intensity value for that spot [12], [13] and [14].

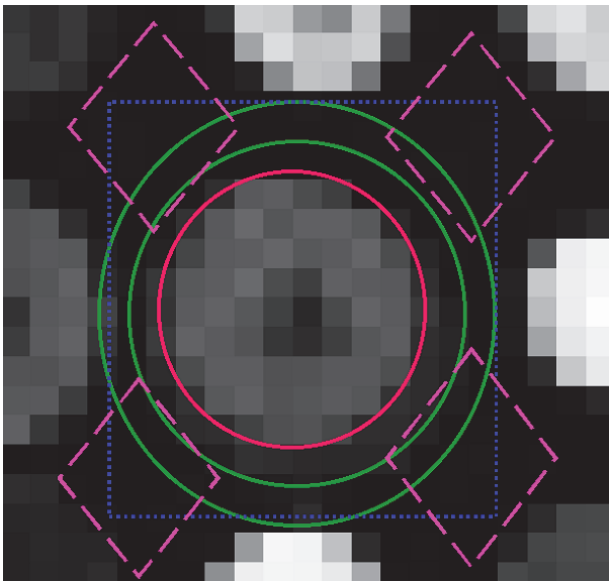

Figure 1. Different background adjustment method.

Table 1 Comparison different background estimation alternatives

FEATURES	(GENEPIX *ET AL.*, 2002)	(EISEN, 1999)	(SUNPAK, 1999)
Method	Genepix	Scanalyze	Quantarry
Background	Four Diamonds	Blue Square & Spot Mask	Two Concentric Circles
BG Calculation	Median	Mean	Median

From the data collected in Table 1, it can be seen that using the median of the pixels' values in the GenePix region decreases the possibility of calculating an erroneous value because the region is very distant from the spot and will not include any pixel that belongs to the spot. Thus, miscalculation is avoided for the bright pixel. However, estimation inaccuracy may occur in the other approaches, e.g., in ScanAlyze where there is overlap between the foreground and background areas, and hence it is inevitable that the values of pixels belonging to the spot will be included in the calculation.

2.2 Most Recent Background Correction Methods

The usual calculation method for finding the corrected value for the intensity of the spot for a two-color microarray is that of subtracting the values of the background intensity of the two colors, red and green (Rb, Gb) from the values of the foreground intensity of the two colors (Rf, Gf). This allows the true spot intensity (R, G) to be more accurately evaluated. This is because the observed foreground intensity value is the sum of the background value and the true intensity value of the spot [15] and [16].

The standard background correction method is usually applied by taking the mean of the foreground values and the median of the background values. The background value is then subtracted from the foreground value to define the true value of the spot intensity. Sometimes the background intensity value is greater than the foreground intensity value, and thus the final result is negative, which is illogical. Therefore, many scientists and researchers have attempted to develop more logical and accurate methods [3].

The Kooperberg correction method is one of the best methods. This method uses the mean of both foreground and background values. Subsequently, a process of numerical integration and convolution is used to calculate the value of the two-color intensity for each spot [1]. This method is restricted to GenePix data and software, and it requires human involvement to function optimally.

A simpler method for avoiding negative intensity values was presented by Edwards using a local median of the background values. This method is similar to the standard method, but when the value after subtraction is less than a specific small threshold value, it undergoes a monotonic function. Whereas when it is larger than the threshold value, it can be considered to be the true intensity of the spot [3] and [17].

Another method for performing the background correction is the morph method. This method uses the main value of the foreground intensity values and morphs the background value, which can be done by performing a morphological opening that involves an erosion operation followed by a dilution operation to find the local maximum and the local minimum. Subsequently, the true value can be estimated by subtracting the background from the foreground value [18]. Similarly, the top-hat method enhances the image to increase the contrast between the spot mask and the background, so that background intensity is decreased automatically [19] and [20].

The last method considered is that of estimating the value of the background intensity is zero. Thus, the true intensity value is taken as the median or mean of the foreground values of the spot. The result of using this method will be a very high value for each spot with both red and green intensities because there is no subtraction involved. However, this method is somewhat illogical since it is not reasonable to assume that all background spot intensities are zero [3] and [19].

Using the data collected in Table 2, different methods of background correction were compared. The standard method is the simplest method but sometimes the result has a negative value when the background intensity is larger than the foreground intensity, which is not meaningful in reality. The Kooperberg integration method decreases the possibility of error, but it is restricted to GenePix's background values. Edwards has also improved the standard method but in a way that is not restricted to any specific software, and furthermore, the threshold value can be changed according to the spot until there are no negative values. The morph estimate is another approach to finding the background value, but it is more liable to error and it is also restricted to spot software. Finally, the method that estimates the background value as zero might possibly serve for the case of an ideal microarray image, but unfortunately, this ideal is rarely encountered in practice.

Table 2 Comparison of different background correction methods

FEATURES	**(YANG *ET AL*., 2002A)**	**(KOOPERBERG *ET AL*., 2002)**	**(EDWARDS, 2013)**	**(MABROUK *ET AL*., 2009)**	**(ALHADIDI *ET AL*, 2007)**	**(BUCKLEY & SPEED, 2001)**
Year	2002	2002	2013	2009	2007	2001
Method	Standard method	Kooperberg	Edwards	Morph	Tophat	No background
Foreground Calculation	mean	mean	median	mean	NA	mean
Background Calculation	median	mean	local median background	image analysis software Spot	NA	Rb = Gb = 0
Spot Calculation	Subtract (Rb & Gb) from (Rf & Gf)	convolution of normal distributions to background	Subtracting and threshold value	Perform morphological opening (erosion & dilation)	mean	Mean of foreground
Advantages and disadvantages	Sometimes: Negative values	the method is restricted to GenePix data	NA	Restricted to spot software	Depending on enhancement of the image	Ignorance of background value

3. MATERIAL AND METHODS

Using MATLAB, we developed a code that can extract the intensity for 10 spots as shown in Figure 2. Using only 10 spots instead of the whole microarray slide makes the process easier and simpler especially when comparing the many different algorithms used. In order to define a method which would be more accurate and suitable for this project, a MATLAB code was developed and tested according to the flowchart in Figure 3.

As can be seen from the flowchart, the MATLAB code starts by importing the whole microarray slide image using the command imread. Then, the image is cropped to 10 spots as shown in Figure 2 using the command imcrop. MATLAB usually reads the image intensity as a matrix of pixels. Thus, our image after cropping is 40*95 pixels although it has only 2*5 spots. Therefore, each spot has a size of around 20-pixel diameters. We then calculated the intensity value for each spot according to the method under consideration.

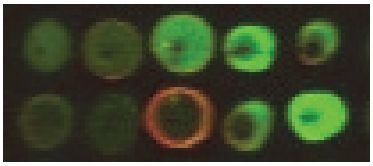

Figure 2. Microarray slide with 10 spot.

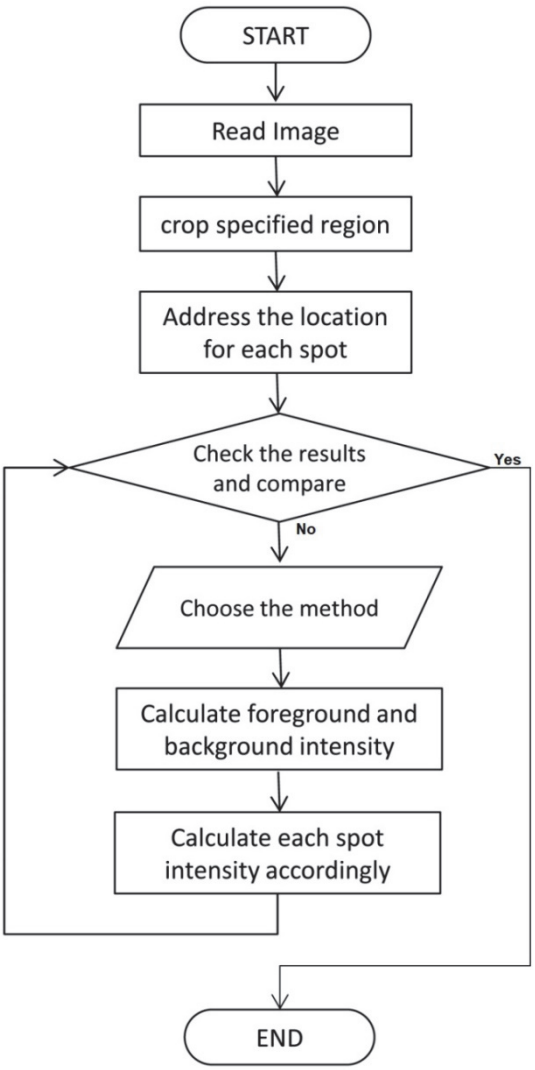

Figure 3. Methodology flowchart.

For the standard method, we first calculated the foreground and background intensities by finding the median value of the pixels inside each area. Then, we calculated the spot intensity by subtracting the background from the foreground. This is similar to the Edwards method, but in the latter, there is a threshold point (value = 0) applied after the subtraction process. However, in the top-hat and no-background methods, the foreground and background intensities were calculated by finding the mean value of the pixels inside each area. We summarized the results for each method in six separate tables for ease of comparison and understanding as shown in section 5.

Finally, one method was chosen and used for extracting the intensities of 100 spots as shown Figure 4. An ideal image with 100 spots was used to check the accuracy of the results. This image is shown in Figure 5.

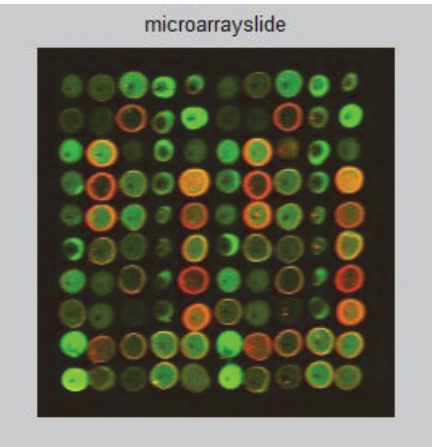

Figure 4. Microarray slide with 100 spots.

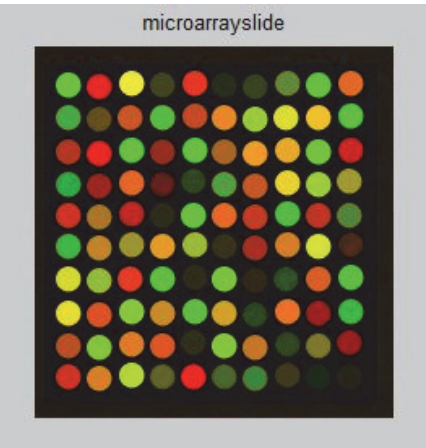

Figure 5. Ideal Microarray slide with 100 spots.

4. RESULTS AND DISCUSSION

Almost all the methods, except the top-hat method, show some similarity, especially when we compare the difference between the green and the red intensities for the same spots. The standard and Edward methods have approximately the same intensity values for the same spots, with only 0.5 differences in values for some spots. The no-background method showed higher intensity values for every spot compared to the other methods. The relationship of the intensity value of the spots in the top-hat method with the same spots in the other methods is variable.

Table 3 shows the results for the standard method, where the result is equal to the median of the background values subtracted from the foreground values, without the use of a threshold. This means that some of the spot values became negative when the background value was greater than the foreground value for a particular pixel. However, since there were a median and many other positive pixels on the spot, the negative value does not appear in the final result of the spot.

The no-background method results are shown in Table 4, with values that are greater than the values of the other methods for each spot. This was due to the fact that there was no subtraction step because the background values for each spot were equal to zero. In addition, not all the intensity values are natural numbers due to the division process during the calculation of the spot intensity where the mean value of the pixels located inside the spot mask is taken.

Spot intensity values for the Edwards method are shown in Table 5; they are almost equal to the standard method except that there are differences of 0.5 of the pixel for some spots. This is due to the use of threshold and median values, since negative numbers will be changed to zero, so that they will not affect the final result when the median is used. However, if we use mean values, the difference will be more than 0.5 of the pixel.
Finally, with regard to the top-hat method in Table 6, the difference between this method and the others is variable due to the image enhancement process using the command imtophat in the MATLAB toolbox. In addition, the difference between the green and the red intensities for the same spot is smaller than in the other methods, while from the image we can see that it is much larger.

Table 3 Results for standard method

STANDARD METHOD		1	2	3	4	5
Red Intensity	Raw 1	4	13	16.5	6.5	2.5
	Raw 2	7.5	2	24	7	20
Green Intensity	Raw 1	40	54.5	103	79.5	12.5
	Raw 2	32	32	30.5	30	130

Table 4 Results for no background method

NO BACKGROUND		1	2	3	4	5
Red Intensity	Raw 1	30.6	34.7	42.6	31.6	32.9
	Raw 2	35	35	76	37.4	38.7
Green Intensity	Raw 1	46.5	37.6	94.1	82.9	43.7
	Raw 2	30.6	34.7	42.6	31.6	32.9

Table 5 Results for edwards method

EDWARDS METHOD		1	2	3	4	5
Red Intensity	Raw 1	4	13	17	6	2.5
	Raw 2	8	2	24	7	20
Green Intensity	Raw 1	40	55	103	80	12.5
	Raw 2	32	32	30	30	130

Table 6 Results for tophat method

TOPHAT METHOD		1	2	3	4	5
Red Intensity	Raw 1	11.92	15.82	25.42	18.8	16.19
	Raw 2	16.09	14.16	55.7	26.17	28.78
Green Intensity	Raw 1	33.86	24.7	81.47	69.37	30.3
	Raw 2	25.14	25.34	31.15	44.52	99.54

From tables 3, 4, 5and 6, it can be clearly seen that almost all the methods can calculate and display the intensities, although the values of the intensities are different due to the different background values. In Table 5, the Edwards method used zero as the threshold value and this method also showed a clear, non-zero background value. Therefore, we propose this method as the most suitable for this project.

To check the results of the Edwards method, an ideal-image microarray slide was used, so that the results were accurate and known as well as they have been settled. Matrix 1 shows the results for

the ideal-image intensities shown in Figure 5. The red and green intensities are displayed in separate matrices with elements arranged in the same way as the arrangement of the spots in the slide, with each element in the matrix representing the intensity of the corresponding spot in the image. Similarly, this method was used to extract the intensities for the real microarray slide shown in Figure 4, as an example of a 100-spot image. Matrix 2 represents the intensities of the image spots.

Red_Intensity =

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

Green_Intensity =

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 Results of Edward background correction method for Ideal Image

Red_Intensity =

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

Green_Intensity =

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 6. Results of Edward background correction method for real Image.

5. CONCLUSION

In this paper, a number of background correction calculations and algorithms were reviewed. The differences and similarities of the existing systems were studied in order to identify the criteria for selecting the best program, the program which gives results closest to the true intensity values. Three methods for allocating and calculating the background intensity values were discussed and compared. These methods were GenePix, ScanAlyze and QuantArry. Five alternative background intensity corrections were applied to a microarray slide image using MATLAB in order to find the most accurate intensity value for each spot in the two-color microarray, and the results compared. Based on the findings, it was concluded that the Edwards method is the best and the most suitable method for correcting background intensity.

ACKNOWLEDGEMENTS

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

REFERENCES

[1] Kooperberg, C., Fazzio, T. G., Delrow, J. J., & Tsukiyama, T. (2002). Improved background correction for spotted DNA microarrays. *Journal of Computational Biology*, 9(1), 55-66.

[2] Lee, M. L. T., Kuo, F. C., Whitmore, G. A., & Sklar, J. (2000). Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. Proceedings of the National Academy of Sciences, 97(18), 9834-9839.

[3] Yang, Y. H., Buckley, M. J., Dudoit, S., & Speed, T. P. (2002). Comparison of methods for image analysis on cDNA microarray data. *Journal of computational and graphical statistics*, 11(1), 108-136.

[4] Mabrouk, M., Fouad, I., & Sharawy, A. (2013). Automated statistical and morphological based gridding methods for noisy microarray image processing. *Journal of Bioinformatics and Intelligent Control*, 2(4), 265-275.

[5] Ahmad MM, Jambek AB, bin Mashor MY. (2013). A study on microarray image gridding techniques for DNA analysis. InElectronic Design (ICED), 2014 2nd International Conference. (pp. 171-175). IEEE.

[6] Sifakis EG, Prentza A, Koutsouris D, Chatziioannou AA. (2012). Evaluating the effect of various background correction methods regarding noise reduction, in two-channel microarray data. Computers in biology and medicine. 2012 Jan 31, 42(1), 19-29.

[7] Smyth GK, Speed T. (2013) Normalization of cDNA microarray data. Methods. 31, 31(4), 265-73.
[8] Geeleher MP, Morris D, Golden A, Hinde J. BioconductorBuntu Users Manual.
[9] Argyropoulos C, Daskalakis A, Nikiforidis GC, Sakellaropoulos GC. (2010). Background adjustment of cDNA microarray images by Maximum Entropy distributions. *Journal of biomedical informatics*, 31, 43(4), 496-509.
[10] Fielden MR, Halgren RG, Dere E, Zacharewski TR. (2002). GP3: GenePix post-processing program for automated analysis of raw microarray data. Bioinformatics, 1, 18(5), 771-3.
[11] Eisen, Michael. (2013). "ScanAlyze user manual." Standford Univ. US.
[12] Luminomics, G. S. I. (2009). "QuantArray Analysis Software, Operator's Manual." GSI Luminomics.
[13] Bengtsson A, Bengtsson H. (2006). Microarray image analysis: background estimation using quantile and morphological filters. BMC bioinformatics, 28, 7(1), 96.
[14] Affymetrix Expression Console Software Version 1.0, User Guide. Santa Clara: Affymetrix Inc.; 2006.
[15] Ritchie ME, Silver J, Oshlack A, Holmes M, Diyagama D, Holloway A, Smyth GK. (2007). A comparison of background correction methods for two-colour microarrays. Bioinformatics, 15, 23(20), 2700-7.
[16] Yin W, Chen T, Zhou SX, Chakraborty A. (2005). Background correction for cDNA microarray images using the TV+ L1 model. Bioinformatics, 15, 21(10), 2410-6.
[17] Edwards D. (2003). Non-linear normalization and background correction in one-channel cDNA microarray studies. Bioinformatics, 1, 19(7), 825-33.
[18] Soille P. (2013). Morphological image analysis: principles and applications. Springer Science & Business Media, 14.
[19] Rao Y, Lee Y, Jarjoura D, Ruppert AS, Liu CG, Hsu JC, Hagan JP. (2008). A comparison of normalization techniques for microRNA microarray data. Stat Appl Genet Mol Biol, 1, 7(1), 22.
[20] Alhadidi B, Zu'bi MH, Suleiman HN. (2007). Mammogram breast cancer image detection using image processing functions. *Information Technology Journal*, 6(2), 217-21.