Methodology article

Profound influence of microarray scanner characteristics on gene expression ratios: analysis and procedure for correction

Heidi Lyng*[1], Azadeh Badiee[2], Debbie H Svendsrud[1], Eivind Hovig[3], Ola Myklebost[3] and Trond Stokke[1]

Address: [1]Department of Biophysics, The Norwegian Radium Hospital, Oslo, Norway, [2]Centre for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway and [3]Department of Tumor Biology, The Norwegian Radium Hospital, Oslo, Norway

Email: Heidi Lyng* - heidi.lyng@labmed.uio.no; Azadeh Badiee - azadeh.badiee@helse-bergen.no; Debbie H Svendsrud - d.h.svendsrud@klinmed.uio.no; Eivind Hovig - eivind.hovig@labmed.uio.no; Ola Myklebost - ola.myklebost@biokjemi.uio.no; Trond Stokke - trond.stokke@labmed.uio.no

* Corresponding author

Abstract

Background: High throughput gene expression data from spotted cDNA microarrays are collected by scanning the signal intensities of the corresponding spots by dedicated fluorescence scanners. The major scanner settings for increasing the spot intensities are the laser power and the voltage of the photomultiplier tube (PMT). It is required that the expression ratios are independent of these settings. We have investigated the relationships between PMT voltage, spot intensities, and expression ratios for different scanners, in order to define an optimal scanning procedure.

Results: All scanners showed a limited intensity range from 200 to 50 000 (mean spot intensity), for which the expression ratios were independent of PMT voltage. This usable intensity range was considerably less than the maximum detection range of the PMTs. The use of spot and background intensities outside this range led to errors in the ratios. The errors at high intensities were caused by saturation of pixel intensities within the spots. An algorithm was developed to correct the intensities of these spots, and, hence, extend the upper limit of the usable intensity range.

Conclusions: It is suggested that the PMT voltage should be increased to avoid intensities of the weakest spots below the usable range, allowing the brightest spots to reach the level of saturation. Subsequently, a second set of images should be acquired with a lower PMT setting such that no pixels are in saturation. Reliable data for spots with saturation in the first set of images can easily be extracted from the second set of images by the use of our algorithm. This procedure would lead to an increase in the accuracy of the data and in the number of data points achieved in each experiment compared to traditional procedures.

Background

Microarray technology is widely used for large scale studies of gene expression levels in cells [1-3]. When spotted cDNA microarrays are used, the expression level is generally measured as a ratio between the fluorescence intensity of two cDNA samples. The samples are labeled with different fluorescent dyes and co-hybridized to an array of DNA probes. The fluorescence intensities are measured by imaging the array in an optical scanner. Many sources of variation are associated with each step of the experimental

procedure [4-6]. Efforts have been made to optimize laboratory protocols and image analysis to increase the accuracy of the results [7-11]. Moreover, non-linear normalization methods have been developed to account for asymmetry in the data [12-14]. Little attention has, however, been paid to the scanning procedure, although it has been proposed that errors in microarray data may be introduced during scanning [12].

The most common scanning configuration utilizes lasers for dye excitation at two different wavelengths, and photomultiplier tubes (PMTs) with two channels for signal detection [15]. A small region of the slide is excited, and light emitted from the dyes is separated from unwanted light using a series of mirrors, filters, and lenses, converted to an electrical signal and digitized. Two images of the entire array, one for each dye, are constructed by moving the slide and/or the lens across the slide in two dimensions. In general, weak signals are detected with lower accuracy than strong ones. An increase in the signal intensities may be achieved by increasing the laser power and/or the PMT voltage, assuming that the expression ratios are independent of these settings.

Increasing the laser power leads to increased intensity of the light emitted from the dyes. The intensity of this light is proportional to number of fluorophores, even at laser intensities high enough to cause dye saturation. In contrast, non-linearity may occur at the detection level. The basic assumption for using PMTs is that the signal from them is proportional to the intensity of the emitted light at a given voltage. However, at low intensities the signals may be influenced by *e.g.* offsets in the electronics. Moreover, saturation of the signals occurs when pixels within the spots reach the upper detection limit of the PMT in intensity. Detailed studies on how and to what extent the PMT voltage influences signal intensities and expression ratios would be useful to ensure that errors in the data are not introduced during scanning.

In the present study we have investigated relationships between PMT voltage, spot intensities, and gene expression ratios for three commercial scanners of two different brands. Our aim was to define an optimal scanning procedure, minimizing the need for non-linear data normalization. We show that reliable data were achieved only for mean spot intensities within the range of 200 to 50 000. To extend the upper limit of this usable range, an algorithm, based on spot intensities within the usable range in two scans, was developed to recalculate the intensities of spots with saturation. Consequently, it was possible to use sufficient voltage during scanning to avoid intensities below the usable range and to retain the spots with saturation. This procedure is useful for all technologies utilizing microarray scanners, regardless of whether cDNA-,

oligo-, protein-, or DNA arrays are considered. The algorithm is supplied through the open-source and free web-based database BioArray Software Environment (BASE) [16], which is released under the GNU General Public License [17].

Results

Relationships between mean spot intensity and PMT voltage are shown in Figure 1 for the three scanners. Three spots selected from the same array are presented; *i.e.*, one spot with saturation at the highest PMT voltages, one weak spot with intensities around 1000, and one spot with intermediate intensities. No correction of saturated intensities, background-subtraction, or normalization was performed. All scanners showed an almost log-linear relationship between intensity and PMT voltage within the intensity range of 200 ($\sim 2^8$) to 50 000 ($\sim 2^{16}$). At a mean spot intensity above 50 000, saturation of pixel intensities occurred, and the mean values increased more slowly with increasing voltage towards the upper limit of 65 536 (2^{16}) of the detection system, for which the intensity of all pixels within the spots was saturated. At higher intensities the values remained constant with increasing voltage. At intensities below 200 the values measured with ScanArray4000 decreased sharply towards zero with decreasing voltage. The intensities measured with GenePix4000B, on the other hand, decreased more slowly, approaching a constant value of about 20. The rate of decrease was different for different spots. Reproducible ratios independent of PMT voltage imply a constant vertical distance on a log-scale between the curves for each of the red and green intensities. This was achieved only within the range of 200 to 50 000; *i.e.*, within the approximate log-linear range of the scanners studied here. Log-linearity is, however, not needed to achieve reproducible ratios, the only requirement is that the vertical distance between the curves is constant on a log-scale for all spots.

The range for reproducible ratios; *i.e.*, the usable intensity range, was the same regardless of dye and array. This is demonstrated for several arrays and ScanArray4000 I in Figure 2, where relationships between intensities in two scans of the same array, obtained at different PMT voltages, are shown for the red (A, B, C) and green (D, E, F) channel separately. The first scan (scan 1) was obtained with the intensities of the weakest spots higher than about 200 and showed saturation in several spots. The second scan (scan 2) was obtained with the brightest spots just below the level of saturation and, hence, with several intensities below 200. Reproducible ratios independent of PMT voltage imply a linear relationship between the intensities in the two scans; *i.e.* a curve with a slope of unity in double logarithmic plots (Fig. 2). The minimum intensity of the spots showing such a relationship was about 200 (scan 2), whereas the maximum intensity was
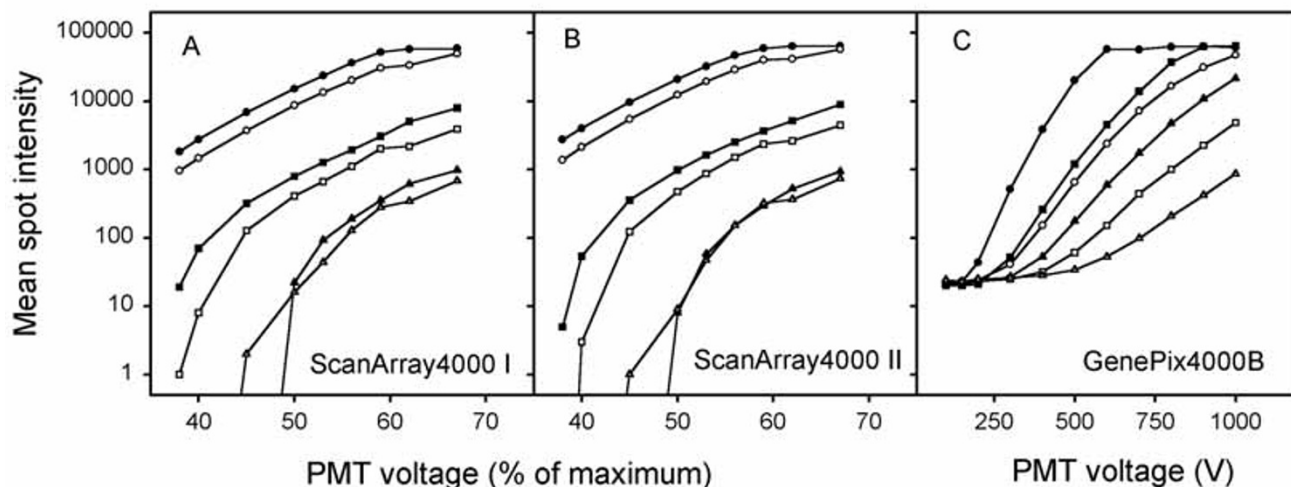
Figure 1
Mean spot intensity *versus* PMT voltage for three scanners of two different brands, ScanArray4000 I (A), ScanArray4000 II (B), and GenePix4000B (C). Three representative spots are presented, one with high intensities (●, ○), one with intermediate intensities (■, □), and one with low intensities (▲, △). The intensities of the green and red channel are represented by closed and open symbols, respectively. The PMT voltage is given in % of maximum voltage for ScanArray4000 and in voltage for GenePix4000B.

about 50 000 (scan 1) regardless of dye or array, as indicated by the dashed lines in Figure 2. The usable intensity range was therefore from 200 to 50 000. At spot intensities above 50 000 (scan 1) and below 200 (scan 2) the relationships deviated from linearity, as expected from the data in Figure 1. The spot intensities above 50 000 contained saturated pixel intensities and were recalculated using the data from scan 2. The corrected data are plotted in Figure 3, where only the high intensity data of Figure 2 are included. The intensity levels of 30 000 and 20 000 in scan 1, which were used in the correction algorithm, are marked together with the level of 50 000. Although spots with mean intensity below 50 000 also could contain some pixels in saturation, only intensities above 50 000 were corrected since this level represented the upper limit of the usable range. The use of our algorithm to correct saturated intensities clearly extended the usable range (Fig. 3). Although this range corresponded to the approximate log-linear range for the scanners studied here, log-linearity is not required for the use of our algorithm.

Plots of normalized expression ratios ($\log_2(Norm \cdot R/G)$, where *Norm* is the total intensity normalization factor) *versus* average intensities ($1/2(\log_2 G + \log_2 R)$) were generated on a double logarithmic scale to visualize the influence of PMT setting and intensity on the expression ratios (Fig. 4). Normalization of microarray data ensures that

the ratios are symmetrically distributed around a value of 1 (2^0), which is a prerequisite for further reliable data analyses [12]. The present data showed such symmetry after normalization whenever the intensities of all spots and background segments were within the intensity range of 200 to 50 000. The use of sufficient PMT voltage to avoid intensities below this range, led to saturation in the brightest spots. This caused asymmetry in the distributions at high intensities, indicating that errors in the expression ratios of these spots had been introduced (Fig. 4A). By using corrected intensities, symmetrical distributions were achieved (Fig. 4B).

Intensity data of one channel, achieved at relatively high PMT voltage, were paired with a series of intensity data of the other channel, achieved at lower PMT voltages, for the array presented in Figure 4A and 4B. All spot intensities above 50 000 were corrected, ensuring that asymmetry in the distributions was caused solely by the scanner behavior at low intensities. The plots demonstrate changes in the expression ratios caused by decreasing the PMT voltage and, hence, the intensities in each of the channels separately. Symmetrical distributions similar to that shown in Figure 4B occurred at voltages higher than 56% (ScanArray4000) or 600 V (GenePix4000B), for which all spot and background intensities in both channels were above 200 (Fig. 4C,4D). The voltage necessary to ensure
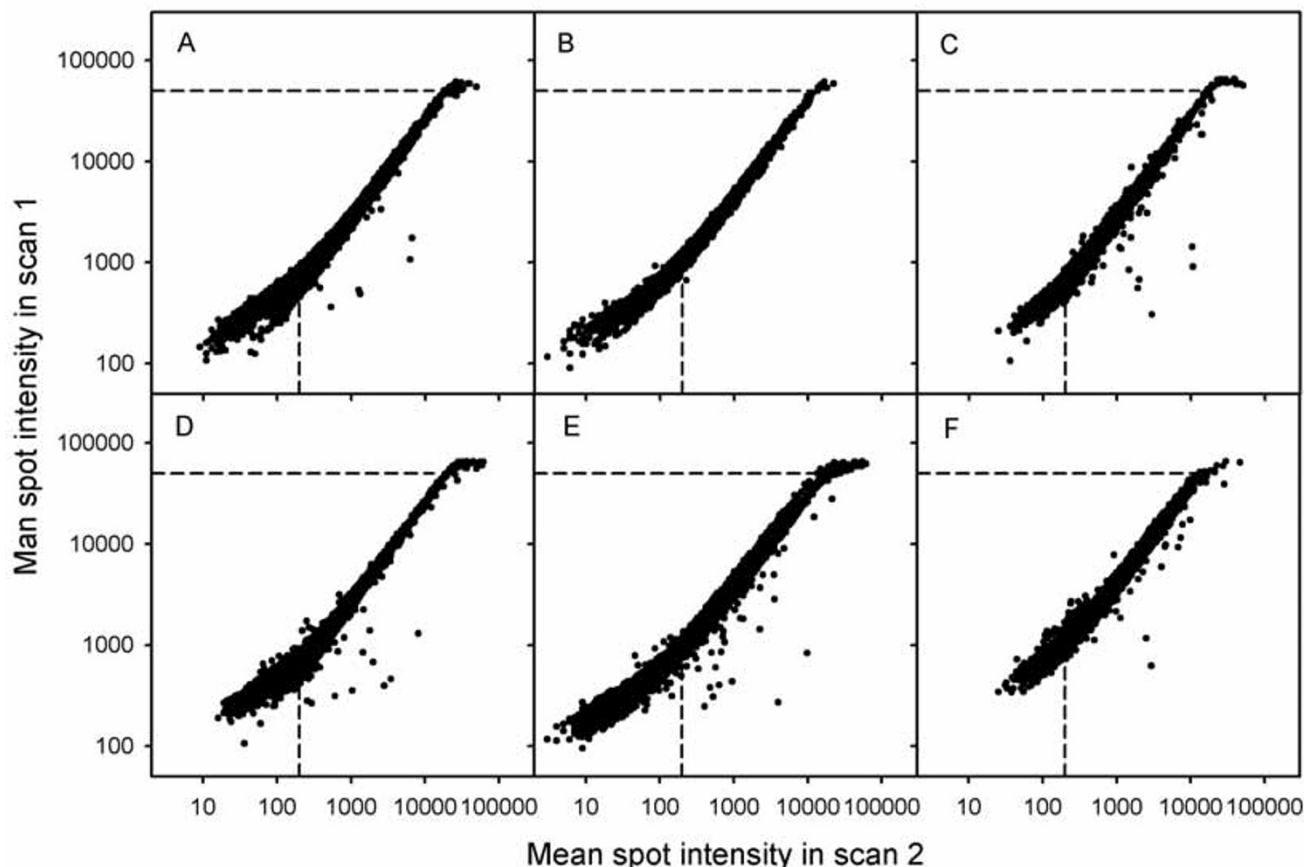
Figure 2
Mean spot intensity in a scan with saturation (scan 1) *versus* mean spot intensities in a scan of the same arrays without satura-
tion (scan 2), obtained at lower PMT voltage. ScanArray4000 I was used during scanning. (A), (B), and (C) Data of the red chan-
nel. (D), (E), and (F) Data of the green channel. (- - - - -) The intensity levels of 50 000 and 200. The PMT voltages (% of
maximum voltage) used in scan 1 and scan 2 were 56% and 51% (A), 59% and 52% (B), 65% and 58% (C), 63% and 56% (D),
65% and 57% (E), 67% and 57% (F).

such intensities differed among arrays, as demonstrated in
Figure 2. The expression ratios achieved at different PMT
voltages were always highly correlated as long as all inten-
sities were above 200 and the saturated intensities were
corrected ($r^2 > 0.98$), consistent with the data presented in
Figure 1.

At lower voltages, spot and background intensities below
200 occurred. The distributions were asymmetrical, show-
ing that considerable errors in the data had been intro-
duced during scanning. The asymmetry was most
pronounced at low intensities. Thus, for ScanArray4000 a
systematic decrease in expression ratios were achieved
when intensities of the red channel decreased below 200
(Fig. 4E,4G,4I), whereas a systematic increase in ratios

was achieved when intensities of the green channel
decreased below this level (Fig. 4F,4H,4J), as compared to
the data in Figure 4B,4C,4D. GenePix4000B, however,
showed the opposite results (data not shown), in accord-
ance with the different relationship between low intensi-
ties and PMT voltage for the two scanner brands (Fig. 1).
Consequently, number of spots with either very high or
very low expression ratio increased considerably,
especially at the lowest voltages (Fig. 4G,4H,4I,4J). When
intensities below 200 occurred in both channels, the dis-
tributions remained fairly symmetrical. However, a signif-
icantly increased spreading at low intensities was
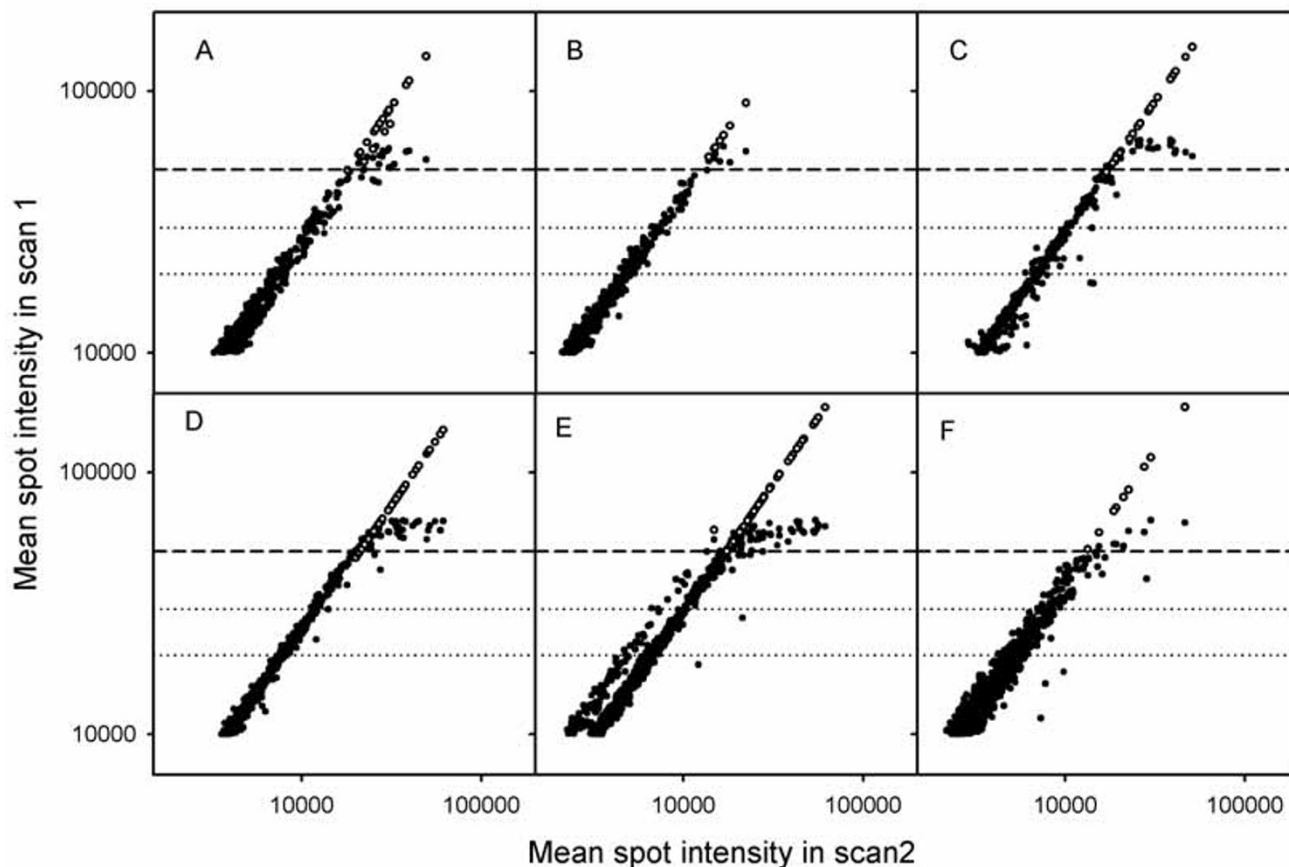observed (data not shown).

Figure 3
Mean spot intensity in a scan with saturation (scan 1) *versus* mean spot intensities in a scan of the same arrays without satura-
tion (scan 2), obtained at lower PMT voltage. The high intensity data of Figure 2 are shown (●) in addition to intensities cor-
rected for saturation (○). ScanArray4000 I was used during scanning. (A), (B), and (C) Data of the red channel. (D), (E), and (F)
Data of the green channel. (- - - -) The intensity level of 50 000. (……) The intensity levels of 30 000 and 20 000. The PMT
voltages (% of maximum voltage) used in scan 1 and scan 2 were 56% and 51% (A), 59% and 52% (B), 65% and 58% (C), 63%
and 56% (D), 65% and 57% (E), 67% and 57% (F).

The expression ratios of a weak and a bright spot in Figure
4 are listed in Table 1 to demonstrate the changes in the
ratios with decreasing voltage. The bright spot had inten-
sities within the range of 200 to 50 000 in all scans,
whereas the intensities of the weak spot decreased below
this range at low voltages. The ratios at low voltages
differed from those achieved when all intensities were
within the range of 200 to 50 000; *i.e.*, at voltages higher
than 50% for both channels using ScanArray4000 or
higher than 500 V (red channel) and 300 V (green chan-
nel) using GenePix4000B, indicating that errors in the
data were induced at the lower voltages. Although the
errors were larger for the weak spot, the ratio of the bright
one also changed (Table 1). The latter change was prima-
rily due to errors in the total intensity normalization fac-
tor caused by unreliably low intensities of the weakest
spots in the array.

Discussion
Significant errors in microarray data can be introduced
during scanning, which was demonstrated in our work.
The errors occurred when the mean spot and background
intensities were either above or below the range of 200 to
50 000 for the scanners used here. Signal saturation at
high intensities is a known phenomenon, and is generally
avoided by reducing the laser power or PMT voltage. The
problems at low intensities, however, are less well known,
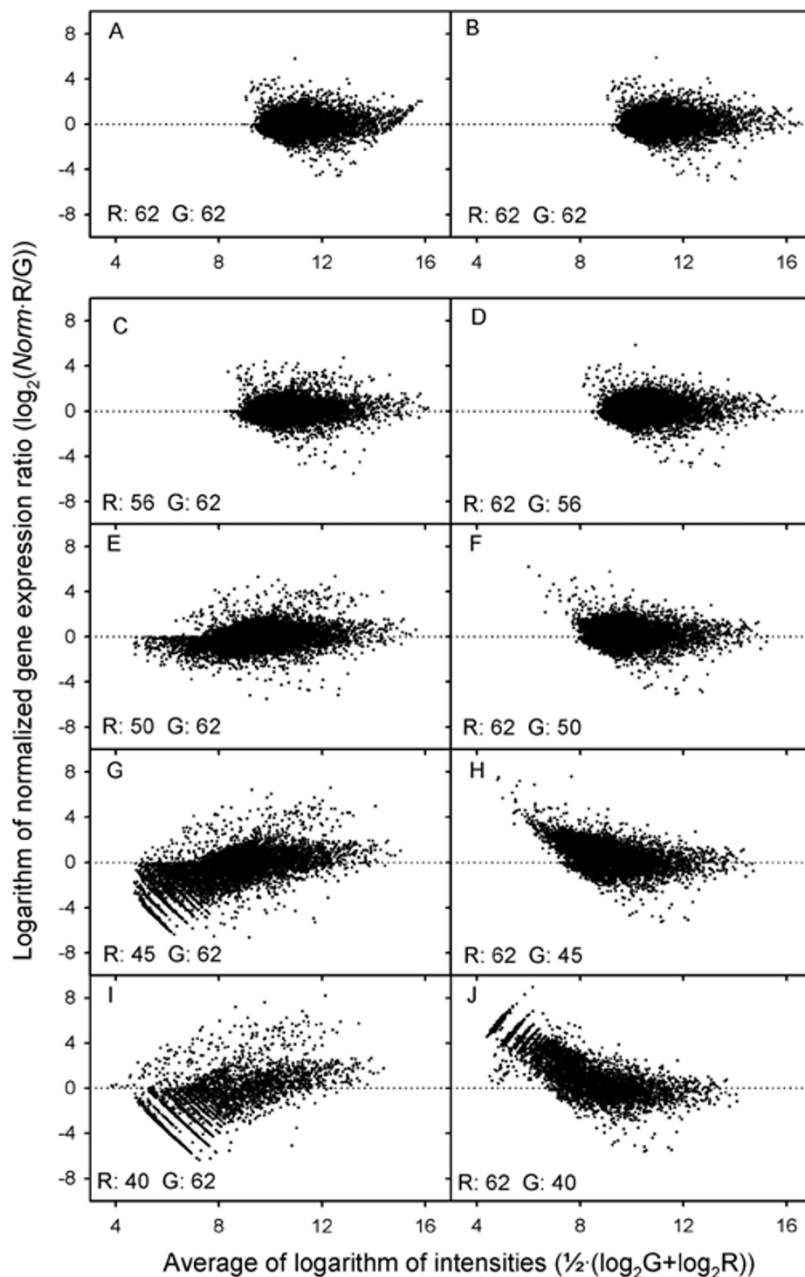and need to be handled differently, as discussed below.

Figure 4
Normalized expression ratio (log$_2$(*Norm*/G), where *Norm* is the total intensity normalization factor) *versus* average intensity in the red (R) and green (G) channel (1/2(log$_2$G+log$_2$R)) plotted on a double logarithmic scale. ScanArray4000 I was used during scanning. The PMT voltage is indicated for the red and green channel separately (% of maximum voltage). The same data set without and with correction of saturated intensities are shown in (A) and (B), respectively. (C), (E), (G), and (I) The intensities in the green channel were above 200 (PMT voltage of 62%), whereas those in the red channel were to an increasing extent below this level (PMT voltages ranging from 56% to 40%). (D), (F), (H), and (J) The intensities in the red channel were above 200 (PMT voltage of 62%), whereas those in the green channel were to an increasing extent below this level (PMT voltages ranging from 56% to 40%). The discontinuity in the distributions at the lowest voltages was due to the apparent discrete values of the lowest intensities and standard deviations, showing values of 1, 2, 3, etc., which occurred because integers were used when assigning the values.

Table 1: Expression ratios of a weak and a bright spot on the same microarray, measured with three scanners at several PMT voltages

	ScanArray4000 I		ScanArray4000 II			GenePix4000B	
PMT (R/G)	Weak spot	Bright spot	Weak spot	Bright spot	PMT (R/G)	Weak spot	Bright spot
62/62*	2.28	0.74	1.88	0.60	800/700*	1.63	0.61
56/62*	2.43	0.76	1.79	0.58	700/700*	1.68	0.55
50/62	2.15	0.74	1.49	0.61	500/700	1.58	0.47
45/62	1.26	0.95	0.43	0.71	400/700	4.98	0.47
40/62	0.55	1.41	0.16	0.76	300/700	14.14	0.98
62/56*	2.25	0.76	1.93	0.62	800/500*	1.71	0.61
62/50	2.36	0.71	2.12	0.58	800/400*	1.79	0.64
62/45	3.37	0.62	3.75	0.50	800/300	1.75	0.62
62/40	92.9	0.48	-	0.41	800/200	0.10	0.46

Data subjected to total intensity normalization are shown. The PMT voltages (PMT) used during scanning are shown for the red (R) and green (G) channel separately as % of maximum voltage (ScanArray4000 I and II) or voltage (GenePix4000B). The PMT voltages leading to intensities within the usable range of the scanners and, hence, reproducible ratios for all spots in the array are indicated (*).

Both scanner brands used here showed a limited intensity range of 200 to 50 000 for which reproducible intensity ratios were achieved independent on PMT voltage. This range was less than the expected range of three orders of magnitude, as specified by the producer of the GenePix4000B scanner [18], and considerably less than the maximum detection range of almost five orders of magnitude for 16 bit scanners. Measurements performed in our institution indicate that the Agilent scanner (Agilent Technologies, Inc., Palo Alto, CA) shows a usable intensity range down to a level of about 100, which is only marginally better than for the scanner brands presented here. Although other scanners may have a larger usable range, the strategies proposed below for reducing the variability in the data are useful for scanners in general.

The reproducibility in intensity ratios at different PMT voltages depends on the intensity *per se*, and not on the voltage. Consequently, different PMT voltage should be used during scanning of different arrays or of different channels in the same array, depending on the intensities of the weakest spots. Moreover, the reproducibility problems are the same regardless of whether the laser power or the PMT voltage is adjusted, since the problems occurs at the detection level, probably in the analog amplifier before the signals are digitized. In principle, the usable intensity range could have been determined by using the laser power, rather than the voltage, to adjust the intensities. In absence of photo bleaching, the highest signal to noise ratio is achieved with maximum laser power, which was therefore used in our study. Significant photo bleaching has, however, occurred with some dye labeled nucleotide batches used at our institutions. To minimize bleaching, the laser power should be reduced. A higher PMT voltage is then needed to achieve spot intensities within the usable range.

The largest asymmetry in the data occurred when the intensities of spots and background segments were below 200. This asymmetry was probably caused by the implementation of an offset voltage in the scanner electronics, preventing weak signals to be dominated by electronic noise. Thus, the use of a negative offset voltage in ScanArray4000 and a positive offset voltage in GenePix4000B may explain why the two scanner brands showed different relationship between the lowest intensities and PMT voltage. Errors in the intensity of the weakest spots and background segments and, hence, in the background-subtracted intensities of all spots and in the total intensity normalization factor occurred. Consequently, all spots showed erroneous expression ratios, although the errors were largest for the weakest spots. The accuracy in the data could therefore only to some extent be improved by excluding low intensity spots. Moreover, the standard deviation of the background intensities was not consistently calculated, leading to errors in the exclusion criteria and/or replacement intensities based on this value. Several studies have shown that the uncertainty in the expression ratio is higher for weak spots than for bright ones [10,19]. This is partially because detection of weak signals during scanning is less reliable than detection of strong ones, even for intensities within the usable range. Our results suggest that the use of spot intensities below this range also may have contributed to these observations.

At intensities above 50 000 erroneous ratios occurred due to saturation of pixels within the spots. The upper limit of the usable range depends on whether median or mean spot intensity is used. The use of median, rather than mean, leads to an upper limit closer to 65 000. Saturation had only minor influence on the total intensity normalization factor and, hence, the expression ratios of the other

spots, as long as the number of spots with saturation was low. However, the expression ratios of the spots with saturation were wrong. Such spots are generally excluded from the analysis or scanning is performed so that no saturation occurs [20,21]. Our algorithm for correcting the intensities of these spots provides a simple method for extending the usable range of the scanners and retaining the spots. The algorithm can easily be implemented in *e.g.* Microsoft Excel or R software based on the formulas listed in the Materials and Methods section. Moreover, an available plug-in module of the same algorithm has been made for BASE in our institution. It was thus possible to link the procedure to a microarray database and analysis platform and make it freely available. The source code of the module is available at http://base.thep.lu.se/plugins[22].

A method for extending the usable range of microarray scanners has been presented previously by Dudley et al. [23]. However, in their work the scanning procedure was not addressed in detail and no specifications of the usable range of different scanners were presented. They used an iterative linear regression algorithm based on intensity data from several scans for correcting spot intensities both above and below the usable range. This method was particularly useful in their experiments, where an extended intensity dynamics of the arrays was achieved due to the use of high intensity oligo samples hybridized in conjunction with cDNAs. In our experiments, two scans obtained at different PMT voltages were always sufficient to cover the range of intensities achieved, making it possible to develop an algorithm that is easier to implement and use.

Other methods, such as spot filtration and non-linear normalization, have already been integrated into microarray analysis to correct for asymmetry in the data. Filtration of spots with intensities below 200 or above 50 000 would reduce the asymmetry. However, the number of spots retained for further analyses is significantly increased by use of our procedure. Moreover, more reliable results are achieved since the background intensities are more accurately determined and the need for non-linear normalization is reduced. Non-linear normalization methods may by themselves introduce uncertainty in the data and lead to diminished ability to detect differentially expressed genes [12].

Conclusions

Improvements in array slide production, sample preparation, and hybridization procedure during the last years have led to increased signal to background ratio and intensity dynamics in microarrays. Our results show that many commercial scanners have a limited usable intensity range that is not suitable for analysis of such high quality arrays. Hence, scanning performed by adjusting the PMT

voltage to avoid saturation may not be the optimal strategy, due to the risk of achieving spot and background intensities below the usable range. A strategy aiming at maximizing the certainty of the weakest spots, by adjusting the voltage to avoid intensities of these spots and background segments below this range, would be more optimal. Since this procedure often leads to a significant number of spots with saturated intensity, methods for extending the usable range are needed. We therefore propose a procedure where two scans are obtained for each of the red and green channel; a primary scan with the lowest intensities within the usable range of the scanner and a secondary scan with the intensities of the brightest spots just below the level of saturation. The primary scan forms the basis of the analysis, whereas the secondary scan is used to correct the intensities of spots with saturation in the primary scan. This procedure will increase the accuracy of the data as well as the number of data points retained in each experiment. However, since the procedure is laborious and time consuming compared to procedures based on a single scan for each channel, revision of the scanner electronics to extend the usable intensity range would be preferable and is highly recommended.

Materials and Methods

Array slides, sample preparation, and hybridization

Array slides produced at the cDNA Microarray Facility at The Norwegian Radium Hospital were used [24]. The arrays contained 13 200 or 15 000 sequence-verified cDNA clones. Total RNA was isolated from dissected tumor tissues (cervical carcinomas) by using Trizol reagent (Life Technologies, Inc., Rockville, MD) and the protocol recommended by the producer. The tumor samples were co-hybridized with another sample from a different region of the same tumor, leading to a limited number of differentially expressed genes with no extremely high or low expression ratios. Sixty to one hundred μg of total RNA was used to produce labeled cDNA by anchored oligo(dT)-primed reverse transcription, using SuperScript II reverse transcriptase (Life Technologies, Inc., Rockville, MD) in presence of either Cy3-dUTP or Cy5-dUTP (Amersham Pharmacia, Piscataway, NJ). The labeled samples were suspended in hybridization buffer and applied to the array slides. Hybridization was performed overnight at 65°C.

Scanning

The slides were imaged at a resolution of 10 μm by use of three commercial PMT based scanners of two brands; two different ScanArray4000 scanners (GSI Lumonics, Wilmington, MA) and a GenePix4000B scanner (Axon Instruments, Inc., Union City, CA). Excitation of Cy3 and Cy5 was performed at a wavelength of 543 nm and 633 nm (ScanArray4000) and 532 nm and 635 nm (GenePix4000B), respectively. A laser power of 100% was

used, since the batches of Cy3- and Cy5-nucleotides employed in our experiments were photo stable. Each slide was subjected to a series of scans with each scanner. The PMT voltage was adjusted simultaneously for the red and green channel in consecutive scans to induce changes in the spot intensities. The signals were digitized into 16 bit/pixel, yielding a maximum detection range from 1 (2^0) to 65536 (2^{16}); *i.e.*, almost five orders of magnitude.

Image analysis

The GenePix 3.0 image analysis software (Axon Instruments, Inc., Union City, CA) was used for spot segmentation and intensity calculation. Bad spots and regions with high unspecific binding of dye were manually flagged and excluded from the analysis. Mean intensity of spot and background segments was derived, and expression ratios were calculated as the ratio between the background-subtracted spot intensities of the red (R) and green (G) channel (R/G). Total intensity normalization of the data was performed as described by Quackenbush [25], and no non-linear normalization methods were applied. When correction of saturated intensities was performed (see below), the corrected intensities were used in the normalization algorithm. Background-subtracted intensities less than two times the standard deviation of the local background were assigned this value to avoid zero or negative values in the ratio calculations [25]. Weak spots with background-subtracted intensity less than this value in both channels were excluded.

Correction of saturated spot intensities

An algorithm was developed to correct saturated spot intensities. The algorithm was based on the requirement of a constant ratio between the intensities of individual spots in two scans, scan 1 and scan 2, achieved at two different PMT voltages. The scans were acquired so that no saturation occurred at the lowest PMT voltage; *i.e.*, in scan 2. Saturation in scan 1 was identified when pixels within the spots reached the upper limit of 65 536 (2^{16}) in intensity. Partially saturated spots occurred when only a fraction of the pixel intensities within the spots were saturated and had a mean value of less than 65 536. In fully saturated spots, all pixels had reached the level of saturation and the mean value was 65 536. Both partially and fully saturated intensities were corrected. The intensities in the two channels were corrected separately.

The corrected value for the saturated spot intensities in scan 1 was calculated as:

$$I_{sat1} = I_{sat2} \cdot F_{cor}$$

where I_{sat1} is the corrected value of the saturated spot intensity, I_{sat2} is the intensity of the corresponding spots in scan 2, and F_{cor} is a correction factor given by:

$$F_{cor} = \frac{1}{m} \sum_{i=1}^{m} \frac{I_{scan1i}}{I_{scan2i}}$$

where *m* is number of spots used in the calculation, I_{scan1} is intensity of selected spots in scan 1, and I_{scan2} is intensity of the corresponding spot in scan 2. It was required that I_{scan1} and I_{scan1} were within the range for reproducible intensity ratios. The spots within the intensity range of 20 000 to 30 000 at the highest PMT voltage were therefore used in these calculations.

Spots with a mean intensity above 50 000 was subjected to the calculations, since these intensities were above the range for which reproducible expression ratios were achieved with the scanners used here (Fig. 1). However, the algorithm can be used to correct possible saturated spot intensities below the level of 50 000 as well.

References

1. Brown PO, Botstein D: **Exploring the new world of the genome with DNA microarrays.** *Nat Genet* 1999, **21:**33-37.
2. Young RA: **Biomedical discovery with DNA arrays.** *Cell* 2000, **102:**9-15.
3. Deyholos MK, Galbraith DW: **High-density microarrays for gene expression analysis.** *Cytometry* 2001, **43:**229-238.
4. Lee ML, Kuo FC, Whitmore GA, Sklar J: **Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations.** *Proc Natl Acad Sci USA* 2000, **97:**9834-9839.
5. Schuchhardt J, Beule D, Malik A, Wolski E, Eickhoff H, Lehrach H, Herzel H: **Normalization strategies for cDNA microarrays.** *Nucleic Acids Res* 2000, **28:**e47.
6. Zhang W, Labordé PM, Coombes KR, Berry DA, Hamilton SR: **Cancer genomics: promises and complexities.** *Clin Cancer Res* 2001, **7:**2159-2167.
7. Wang X, Ghosh S, Guo S-W: **Quantitative quality control in microarray image processing and data acquisition.** *Nucleic Acid Res* 2001, **29:**e75.
8. Wildsmith SE, Archer GE, Winkley AJ, Lane PW, Bugelski PJ: **Maximization of signal derived from cDNA microarrays.** *BioTechniques* 2001, **30:**202-208.
9. Yang YH, Buckley MJ, Speed TP: **Analysis of cDNA microarray images.** *Brief Bioinform* 2001, **2:**341-349.
10. Yue H, Eastman PS, Wang BB, Minor J, Doctolero MH, Nuttall RL, Stack R, Becker JW, Montgomery JR, Vainer M, Johnston R: **An evaluation of the performance of cDNA microarrays for detecting changes in global mRNA expression.** *Nucleic Acids Res* 2001, **29:**e41.
11. Lage JM, Hamann S, Gribanov O, Leamon JH, Pejovic T, Lizardi PM: **Microgel assessment of nucleic acid integrity and labeling quality in microarray experiments.** *BioTechniques* 2002, **32:**312-314.
12. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP: **Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation.** *Nucleic Acids Res* 2002, **30:**e15.

13. Kepler TB, Crosby L, Morgan KT: **Normalization and anlysis of DNA microarray data by self-consistency and local regression.** *Genome Biol* 2002, **3:**research0037.1-0037.12.
14. Workman C, Jensen LJ, Jarmer H, Berka R, Gautier L, Nielsen HB, Saxild H-H, Nielsen C, Brunak S, Knudsen S: **A new non-linear normalization method for reducing variability in DNA microarray experiments.** *Genome Biol* 2002, **3:**research0048.1-0048.
15. Schermer MJ: **Confocal scanning microscopy in microarray detection.** In: *DNA microarrays – a practical approach Volume 1*. Edited by: *Schena M. Oxford: Oxford University Press*; 1999:17-42.
16. Saal LH, Troein C, Vallon-Christersson J, Gruvberger S, Borg Å, Peterson C: **BioArray Software Environment (BASE): A platform for comprehensive management and analysis of microarray data.** *Genome Biol* 2002, **3:**software0003.1-0003.6 [http://base.thep.lu.se].
17. **GNU General Public License** [http://www.gnu.no]
18. **Axon Instruments** [http://www.axon.com/GN_GenePix4000.html]
19. Jenssen T-K, Langaas M, Kuo WP, Smith-Sørensen B, Myklebost O, Hovig E: **Analysis of repeatability in spotted cDNA microarrays.** *Nucleic Acid Res* 2002, **30:**3235-3244.
20. Hsiao LL, Jensen RV, Yoshida T, Clark KE, Blumenstock JE, Gullans SR: **Correcting for signal saturation errors in the analysis of microarray data.** *BioTechniques* 2002, **32:**330-336.
21. Smyth GK, Yang YH, Speed T: **Statistical issues in cDNA microarray data analysis.** 2002 [http://www.stat.berkeley.edu/users/terry/zarray/Html/papersindex.html].
22. . Standard distribution of BASE - http://base.thep.lu.se/plugins. Plug in avaliable also from Vegard Nygaard (vegard.nygaard@klinmed.uio.no)
23. Dudley AM, Aach J, Steffen MA, Church GM: **Measuring absolute expression with microarrays with a calibrated reference sample and an extended signal intensity range.** *Proc Natl Acad Sci USA* 2002, **99:**7554-7559.
24. **The Norwegian Microarray Consortium** [http://www.med.uio.no/dnr/microarray/english.html]
25. Quackenbush J: **Microarray data normalization and transformation.** *Nat Genet* 2002, **32:**496-501.