

Lawrence Berkeley National Laboratory

Lawrence Berkeley National Laboratory

Title

Empirical Evaluation of a New Method for Calculating Signal to Noise Ratio (SNR) for Microarray Data Analysis

Permalink

<https://escholarship.org/uc/item/9r3512mc>

Author

He, Zhili

Publication Date

2008-07-14

1 **ABSTRACT**

2

3 Signal-to-noise-ratio (SNR) thresholds for microarray data analysis were experimentally
4 determined with an oligonucleotide array that contained perfect match (PM) and mismatch (MM)
5 probes based upon four genes from *Shewanella oneidensis* MR-1. A new SNR calculation, called
6 signal to both standard deviations ratio (SSDR) was developed, and evaluated along with other
7 two methods, signal to standard deviation ratio (SSR), and signal to background ratio (SBR). At
8 a low stringency, the thresholds of SSR, SBR, and SSDR were 2.5, 1.60 and 0.80 with
9 oligonucleotide and PCR amplicon as target templates, and 2.0, 1.60 and 0.70 with genomic
10 DNA as target templates. Slightly higher thresholds were obtained at the high stringency
11 condition. The thresholds of SSR and SSDR decreased with an increase in the complexity of
12 targets (e.g. target types), and the presence of background DNA, and a decrease in the
13 composition of targets, while SBR remained unchanged under all situations. The lowest
14 percentage of false positives (FP) and false negatives (FN) was observed with the SSDR
15 calculation method, suggesting that it may be a better SNR calculation for more accurate
16 determination of SNR thresholds. Positive spots identified by SNR thresholds were verified by
17 the Student t-test, and consistent results were observed. This study provides general guidance for
18 users to select appropriate SNR thresholds for different samples under different hybridization
19 conditions.

1 INTRODUCTION

2 Microarrays have become a routine tool for studying gene functions, regulations and networks in
3 a variety of biological systems. Currently, this technology has been also applied to drug
4 discovery and validation (7), microbial diagnostics (4, 10, 16, 20, 22, 31), mutation and single
5 polymorphism nucleotide (SNP) detection (9), strain comparison and genotyping (1, 8, 21), species
6 identification (32), array sequencing (35), environmental detection and monitoring (5, 6, 13, 24,
7 27, 28, 33), and evolutionary processes (14). However, due to small spot sizes, different degrees
8 of uniformity of printing pins, and uneven hybridization, microarray spots inherently have
9 relatively high noise, which presents a variety of challenges for quantitative analysis of
10 microarray data. For example, how to distinguish a real signal from its background is still an
11 unsolved problem, and a subset of this question is what parameters and thresholds should be used
12 to differentiate a signal from a noise.

13 The signal-to-noise ratio (SNR) has been used to define a positive spot, and two general
14 methods are currently used to calculate SNR values. One is to use the ratio of the differences
15 between signal mean and background noise divided by background standard deviation (2). This
16 calculation method has been commonly used in many signal-processing disciplines, such as
17 radio, electronics and imaging (2, 30), and the threshold is usually set to 3.0 (30). The other
18 method is to use the ratio of signal median divided by background median, and the threshold was
19 set to 1.50 (26), and it was modified to calculate the SNR for a probe with replicate spots and set
20 the threshold of 2.0 (18, 19). However, the determination of these thresholds is arbitrary and has
21 not been experimentally validated. Although the background standard deviation of pixel
22 intensities for each spot is included in the first calculation method, the signal standard deviation
23 is not considered in either of the two SNR calculation methods. In addition, an SNR threshold

1 may vary with different types of targets, target compositions, and hybridization conditions, and
2 hence it could be difficult to set a universal SNR threshold. Therefore new SNR calculation
3 methods to include both signal and background standard deviations and experimental evaluations
4 of SNR thresholds are needed.

5 The objectives of this study were to: (i) evaluate a new SNR calculation method for SNR
6 calculation, (ii) determine appropriate SNR thresholds for differentiating signals from noises
7 based on different SNR calculation methods, and (iii) examine the effects of target types,
8 background DNA, and target compositions on the threshold determination. Our results
9 demonstrated that our new calculation performed better than two other existing calculations, and
10 that SNR thresholds were affected by hybridization stringency, types of target templates,
11 background DNAs, and compositions of the target templates. Those results provide general
12 guidance for users to select appropriate SNR thresholds under different conditions.

13

14 **METHODS AND MATERIALS**

15 **Oligonucleotide probe design and microarray construction**

16 50mer and 70mer perfect match (PM) and mismatch (MM) oligonucleotide probes were prepared
17 as previously described (12). Briefly, four genes (*SO1679*, *SO1744*, *SO2680*, and *SO0848*) were
18 selected from the *Shewanella oneidensis* MR-1 genome. For each gene, one 50- or 70-mer PM
19 probe, and 45 MM (with 1 to 37 mismatches) probes were generated with three random MM
20 probes at each level. All 368 designed oligonucleotides were commercially synthesized without
21 modification by MWG Biotech Inc. (High Point, NC). The concentration of oligonucleotide
22 probes was adjusted to 100 pmol/μl. Oligonucleotide probes prepared in 50% DMSO (Sigma
23 Chemical Co., MO) were spotted onto UltraGAPS glass slides (Corning Life Science, NY) using

1 a PixSys 5500 robotic printer (Cartesian Technologies Inc., CA). Each probe had four replicates
2 on a single slide. In total, there were 1472 (368 x 4) spots on the array. After printing, the
3 oligonucleotide probes were fixed onto the slides by UV cross-linking (600 mJ of energy)
4 according to the protocol of the manufacturer (Corning Life Science, NY).

5 **Target template preparations**

6 Four 70mer artificial targets (T1-SO1679, T2-SO1744, T3-SO2680 and T4-SO0848) that were
7 complementary to the 70mer PM probes were synthesized by the Molecular Structure Facility at
8 Michigan State University (East Lansing, MI). The artificial oligonucleotide targets were labeled at
9 the 5'-end with Cy5 (T1-SO1679, T2-SO1744 and T3-SO2680) or Cy3 (T4-SO0848) fluorescent
10 dye during synthesis. The 70mer oligonucleotide targets also contained sequences of 50mer
11 oligonucleotide targets.

12 Gene-specific primers were selected for the four selected genes ([Supplementary Table S1](#))
13 with each PCR product about 500 bp covering both 50mer and 70mer probe sequences. Each
14 gene was amplified with *S. oneidensis* MR-1 genomic DNA as template using the standard PCR
15 amplification protocol. The amplified PCR products were purified using the QIAquick PCR
16 Purification Kit (QIAGEN Inc., CA) according to the protocol of the manufacturer. The purified
17 PCR fragments were visualized and checked the sizes via agarose gel electrophoresis, and then
18 quantified using the PicoGreen dsDNA Assay Kit (Invitrogen, CA).

19 Genomic DNAs from four bacteria were also used as target DNAs. *Shewanella oneidensis*
20 MR-1, *Escherichia coli* S17, and *Pseudomonas sp.* strain G179 were grown in the LB medium to
21 stationary phases, and *Desulfovibrio vulgaris* Hildenborough were grown in the LS medium. The
22 cells were collected by centrifugation at 4000 x g at room temperature for 10 min. Their genomic
23 DNAs were isolated and purified as described previously ([34](#)). *Methanococcus maripludis*

1 gDNA was provided by Sergey Stolyar at the University of Washington (Settle, WA). The yeast
2 *Saccharomyces cerevisiae* was grown in the YPD medium to the saturation, and its genomic
3 DNA was extracted using the glass bead method as described by Hoffman and Winston (15).

4 To test how bacterial ratios affect the determination of SNRs, *S. oneidensis* MR-1 gDNA
5 was mixed with other four bacterial gDNAs (*D. vulgaris* Hildenborough, *E. coli* S17,
6 *Pseudomonas sp.* strain G179, and *M. maripludis*) with three different ratios: A = 10 (*S.*
7 *oneidensis* MR-1):1:1:1:1; B = 1 (*S. oneidensis* MR-1):1:1:1:1, and C = 1 (*S. oneidensis* MR-
8 1):10:10:10:10, respectively. Each sample had the same amount of total gDNA (2.5 µg).

9 **Probe labeling, microarray hybridization, and image quantification**

10 PCR amplicons, the purified genomic DNA from pure cultures (500 ng), and mixed genomic
11 DNAs (2.5 µg) were fluorescently labeled by random priming using Klenow fragment of DNA
12 polymerase (12). Mixture I (35 µl) containing certain amounts (as indicated in different
13 experiments) of genomic DNA and 20 µl of random primers (Invitrogen, CA) was heated at
14 98°C for 3 to 5 min, cooled on ice and then centrifuged. Mixture II (15 µl) containing 1µl of 5
15 mM dATP, dGTP and dTTP and 2.5mM dCTP, 2 µl (80 U) of Klenow (Invitrogen, CA) and 0.5
16 µl of Cy3 dye (Amersham BioSciences, UK) was added to mixture I. A total of 50 µl labeling
17 reaction solution was incubated for 3 hr at 42°C. The labeling reaction was terminated by heating
18 at 98°C for 3 min. The tubes were removed and placed on ice. The labeled cDNA targets were
19 purified immediately using a QIAquick PCR purification column and concentrated in a Savant
20 Speedvac centrifuge (Savant Instruments Inc., Holbrook, NY).

21 Labeled PCR amplicons, or genomic DNAs were resuspended in a 25 µl of hybridization
22 solution containing 50% formamide, 5 x saline-sodium citrate (SSC), 0.1% SDS, and 0.1 mg/ml
23 of herring sperm DNA (Invitrogen, CA). The hybridization solution was incubated at 95-98°C

1 for 5 min, centrifuged to collect condensation, and kept at 50°C. The solution was immediately
2 applied onto the microarray slide and hybridization was carried out in a waterproof Corning
3 Hybridization chamber (Corning Life Science, NY) submerged in a 45°C water bath in the dark
4 for 16 h (12). Washing was performed immediately in the following steps: (i) in a solution
5 containing 2 x SSC and 0.1% SDS at 40°C for 5 min and repeated once, (ii) in a solution
6 containing 0.1 x SSC and 0.1% SDS at room temperature for 10 min and repeated once, and (iii)
7 in 0.1 x SSC at room temperature for 2 min and repeated once. Slides were dried by compressed
8 air prior to scanning. The same batch slides and the same settings were used for all experiments.
9 The laser power was set to 95%, and photomultiplier tube (PMT) efficiency was set to 70%. Five
10 slides (4 replicated spots in each slide) were used for each condition, and hence each spot had up
11 to 20 data points. Hybridized microarray slides were scanned using ScanArrayTM Express
12 microarray analysis system (Perkin Elmer®, MA). Spot signal, spot quality, and background
13 fluorescent intensities of scanned images were quantified with ImaGene version 6.0
14 (Biodiscovery Inc., Los Angeles, CA).

15 **Data analysis**

16 Data analysis included the following four major steps. (i) Defining positive and negative spot
17 pools: Microarray detection mainly depends on probe specificity and hybridization stringency
18 (e.g. temperature), and two levels of stringency were used in this study. A high-level stringency
19 is expected to eliminate cross-hybridization for the probes with a higher probe-target similarity, a
20 longer continuous stretch length, and a lower free energy. At both stringencies, positive and
21 negative pools were defined (Supplementary Table S2 & S3). At the high stringency, a positive
22 50mer probe has a sequence identity >90%, a stretch length >20, and free energy <-35 kcal/mol
23 with its non-targets, and a negative probe has a sequence identity ≤90%, a stretch length ≤20,

1 and free energy ≥ -35 kcal/mol with its non-targets. Our previous experimental results showed
2 that such high stringency hybridization can be achieved at 50°C plus 50% formamide (17).
3 Similarly, a positive 70mer probe has a sequence identity $>90\%$, a stretch length >25 , and free
4 energy < -50 kcal/mol with its non-targets, and a negative probe has a sequence identity $\leq 90\%$, a
5 stretch length ≤ 25 , and free energy ≥ -50 kcal/mol with its non-targets. At a low stringency, a
6 positive 50mer probe has a sequence identity $>85\%$, a stretch length >15 , and free energy < -30
7 kcal/mol with its non-targets, and a negative probe has a sequence identity $\leq 85\%$, a stretch
8 length ≤ 15 , and free energy ≥ -30 kcal/mol with its non-targets (12). The low stringency generally
9 corresponds to hybridization at 42°C plus 50% formamide. Similarly, a positive 70mer probe has
10 a sequence identity $>85\%$, a stretch length >20 , and free energy < -40 kcal/mol with its non-
11 targets, and a negative probe has a sequence identity $\leq 85\%$, a stretch length ≤ 20 , and free energy
12 ≥ -40 kcal/mol with its non-targets (12). In addition, the probes that do not qualify for either
13 positive pool or negative pool were ignored for further analysis. (ii) Microarray spot analysis:
14 Spot intensity data were extracted from ImaGene output files. The values for gene ID, flag,
15 signal mean (\bar{S}), background mean (\bar{B}), signal standard deviation (σ_s), and background
16 standard deviation (σ_b) were extracted from ImaGene output files. After the removal of bad
17 spots, the rest of spots (including potential empty spots and good spots) were kept for further
18 analysis. All processes were conducted with Microsoft Excel software. (iii) Calculation of SNR
19 values: For each spot, three methods were used to calculate SNR values:

$$20 \quad SSR = \frac{(\bar{S} - \bar{B})}{\sigma_b} \quad (\text{I})$$

$$21 \quad SBR = \frac{\bar{S}}{\bar{B}} \quad (\text{II})$$

1
$$SSDR = \frac{(\bar{S} - \bar{B})}{(\sigma_s + \sigma_b)} \quad (\text{III})$$

2 Where \bar{S} and \bar{B} are the signal mean and the background mean of pixel intensities, respectively,
3 and σ_s and σ_b are the standard deviation of signal and background, respectively. Based on false
4 positive and false negative spots at different values of SSR, SBR, and SSDR (in comparison with
5 the defined positive and negative spot pools), their thresholds were determined by (a) minimizing
6 false positives, (b) minimizing false negatives, and (c) optimizing the overall percentage of false
7 positives and false negatives. (vi) The student-t test analysis of threshold-identified positive
8 spots: The values of signal (S) and background (B) were extracted for a probe with replicate
9 spots from ImaGene output files, and their mean (\bar{S}_m and \bar{B}_m , respectively) and standard
10 deviation ($\sigma_{s,m}$ and $\sigma_{b,m}$, respectively) values were calculated. Outliers were removed if $|S - \bar{S}_m|$
11 $\geq 2.0 * \sigma_{s,m}$, or $|B - \bar{B}_m| \geq 2.0 * \sigma_{b,m}$, and this process continued until outliers were recursively
12 removed. The final \bar{S}_m , \bar{B}_m , $\sigma_{s,m}$ and $\sigma_{b,m}$ were used for the Student t-test, and the significance
13 between \bar{S}_m and \bar{B}_m was statistically evaluated for each probe at a given p value.

14 **Data analysis for *Desulfovibrio vulgaris* Hildenborough microarrays**

15 Both wild type and Δfur mutant of *D. vulgaris* cells were grown in the LS4D medium with 60
16 μM of iron, and microarray data were obtained as previously described (3). SSDR method was
17 used to detect positive spots with the threshold of 0.80, and details of data analysis were
18 conducted as previously described (3).

19 **Data analysis for GeoChip with a soil sample**

20 A soil sample was taken from a plot at BioCON (23), and 5 g of soil was used to extract DNA.
21 GeoChip (13) was used to detect functional genes in such a microbial soil community. SSR, SBR
22 and SSDR were used to detect positive spots with thresholds of 2.0, 1.6, and 0.8, respectively,

1 and details for labeling, hybridization, and scanning were performed as described previously
2 (13).

3

4 **RESULTS**

5 **A new SNR calculation method**

6 To consider the signal intensity and background noise as well as their standard deviations for
7 each spot, a new calculation method, termed SSDR (signal to both standard deviations ratio),
8 was developed. SSDR differs from other two SNR calculation methods (SSR and SBR) in that it
9 takes account into the signal standard deviation as a part of the denominator. The relationship
10 between SSDR and signal or background intensity (together with their standard deviations) can
11 be simply represented in Fig. 1, which shows both signal and background standard deviations are
12 equally important for determination of SNR thresholds. When SSDR is ≥ 1.0 , the difference
13 between the signal intensity and background noise is equal or larger than the sum of the signal
14 and background standard deviations. In this case, the pixel values of signal intensity are
15 completely separated from those of background noises (Fig. 1). Intuitively, such a spot should
16 represent positive signal. When SSDR < 1.0 , overlaps of the pixel values between signals and
17 background noises exist (Fig. 1). In this case, some spots could be positive while some are not,
18 but the key question is what is the minimum SNR (e.g. SSDR) threshold for distinguishing the
19 signal from its background noise. Thus in this study, we will experimentally determine the
20 threshold of SSDR for differentiating signals from noises.

21 **Experimental determination of SNR thresholds**

22 To determine appropriate thresholds for distinguishing signal from noise for a single spot on the
23 array, four synthesized targets were hybridized with the array at a final concentration of 10 pg

1 per oligonucleotide. Based on the predefined positive and negative pools at the low stringency,
2 60 (27 for 50mer, 33 for 70mer) probes are expected to be positives, 249 negative, and 59
3 ignored (Supplementary Table S2). The ignored probes fail to satisfy the definition of positive or
4 negative spots. Based on the predicted pools of the positive and negative spots, the number of the
5 false positive (FP) and false negative (FN) spots were calculated at different scenarios. First,
6 false positive spots were minimized. To have no false positives, the thresholds of SSR, SBR, and
7 SSDR should be 5.0, 5.0, and 1.0, respectively (Table 1, Fig. 2). If 1% FP spots were allowed,
8 the thresholds were 4.0 for SSR, 3.5 for SBR, and 0.90 for SSDR (Table 1, Fig. 2). The
9 thresholds would be 2.0, 1.8, and 0.70 for SSR, SBR and SSDR, respectively when 5% FP spots
10 could be tolerated (Table 1, Fig. 2). Second, false negatives were minimized. The thresholds of
11 SSR, SBR, and SSDR should be 0.5, 0.5, and 0.3, respectively if there were no FN spots (Table
12 1, Fig. 2). If 1% FN spots were allowed, the thresholds were 1.5 for SSR, 1.2 for SBR, and 0.70
13 for SSDR (Table 1, Fig. 2). The thresholds would be 2.5, 1.6, and 0.85 for SSR, SBR and SSDR,
14 respectively when 5% FN spots were allowed (Table 1, Fig. 2). In addition, the thresholds of
15 SSR, SBR and SSDR were determined by optimizing the total percentage of FP and FN spots.
16 Generally speaking, higher percentages of FP were observed at a lower threshold of SSR, SBR,
17 or SSDR. For example, the percentages of FP were 11.8%, 12.2%, and 7.9% at SSR = 1.5, SBR
18 = 1.4, and SSDR = 0.5, respectively, which led to 13.0%, 14.9%, and 8.3% of total percentages
19 of FP and FN spots, respectively (Fig. 2). On the contrary, higher percentages of FN were
20 observed at a higher threshold of SSR, SBR, or SSDR. For example, the percentages of FN were
21 17.1%, 19.0%, and 12.8% at SSR = 4.0, SBR = 4.0, and SSDR = 1.2, respectively, resulting in
22 18.2%, 19.9%, and 13.1% of total percentages of FP and FN spots, respectively (Fig. 2).
23 However, relatively low and stable percentages of FP and FN spots were shown when the values

1 of SSR, SBR or SSDR were in a certain range. For example, when SSR were between 2.0 and
2 3.0, the percentages of FP and FN spots were 8.0-9.7%; those percentages were 10.0-14.9%
3 when SBR were 1.4-3.0; SSDR were 0.6-1.0 when those percentages were 5.0-8.0% (Fig. 2).
4 Therefore, the above results indicate that the thresholds of SSR, SBR and SSDR can be in a
5 certain range with a relatively low percentage of FP and FN spots although optimal thresholds
6 were determined to be $SSR = 2.5$, $SBR = 1.6$, and $SSDR = 0.80$.

7 Under a high stringency, 33 (13 for 50mer and 20 for 70mer) probes were positives, 280
8 (147 for 50mer and 137 for 70mer) were negative, and 55 were ignored (Supplementary Table
9 S3). The thresholds of SSR, SBR and SSDR were determined using the same strategies as
10 described above. First, through the minimization of false positives, the thresholds of SSR, SBR,
11 and SSDR were determined to be 5.0, 5.0 and 1.1, respectively when no FP spots were allowed;
12 those thresholds were 4.5 for SSR, 4.0 for SBR and 1.0 for SSDR if 1% FP spots were allowed;
13 if 5% FP spots were tolerated, those thresholds of SSR, SBR and SSDR were 2.5, 2.0 and 0.70,
14 respectively (Table 1, Fig. 3). Second, through the minimization of false negatives, the
15 thresholds of SSR, SBR and SSDR were determined to be approximately 1.0, 1.0, and 0.5,
16 respectively when no FN spots were allowed; if 1% FN spots were allowed, those thresholds
17 were 2.0 for SSR, 1.4 for SBR, and 0.75 for SSDR; they would be 3.0 for SSR, 1.8 for SBR, and
18 0.95 for SSDR if 5% FN spots were tolerated (Table 1, Fig. 3). Finally, by optimizing the total
19 percentage of FP and FN spots on the array, the thresholds of SSR, SBR and SSDR were
20 determined to be 3.0, 2.0 and 0.90, respectively (Fig. 3). The results demonstrated that the
21 thresholds of SSR, SBR and SSDR increased with an increase in stringency of defined positive
22 and negative probe pools. In addition, both Fig. 2 and Fig. 3 showed that the lowest percentages
23 of FP and FN spots were observed with the SSDR calculation, and that an optimization of

1 percentage of FP and FN appeared to be the best method for SNR determination. Therefore, for
2 further experiments, the defined positive and negative pools with the low stringency were used,
3 and an optimization of false positives and false negatives was considered the best method for
4 SNR determination.

5 **Effects of target types on the SNR threshold determination**

6 To determine the impacts of target types on the threshold selection, 100 pg of each PCR
7 amplicon or 500 ng of *S. oneidensis* MR-1 gDNA was also labeled with Cy3 and hybridized with
8 the array, and the thresholds of SNR, SBR and SDR were determined by optimizing the
9 percentage of FN and FP spots. The same thresholds were obtained for PCR amplicon targets as
10 the synthesized oligonucleotides although the PCR amplicon targets caused slightly higher
11 percentages of total FN and FP than synthesized oligonucleotides. For example, the thresholds of
12 SDR were 2.5 for oligonucleotide and PCR amplicon targets when the percentages of FP and FN
13 were 8.0% and 8.7%, respectively (Fig. 4A). However, the thresholds of SDR of 2.0 (Fig. 4A)
14 and SDR of 0.70 (Fig. 4C) for gDNA were lower than those for synthesized oligonucleotides,
15 or PCR amplicons. The percentages of total FN and FP of gDNA were a bit higher than
16 synthesized oligonucleotide, or PCR amplicon targets (Fig. 4). For example, the percentage of
17 FN and FP was 7.1% for gDNA compared to 5.0% for oligonucleotide targets and 6.51% for
18 PCR targets when the SDR thresholds of 0.8, 0.8 and 0.7 were used for oligonucleotide, PCR
19 amplicon, and gDNA targets, respectively (Fig. 4C). In contrast to SDR and SDR, SBR
20 remained unchanged with different types of targets. The results also confirmed that the lowest
21 percentage of false positives and false negatives was observed with the SDR calculation
22 method.

23 **Effects of background DNA on threshold determination**

1 When microarrays are used for community analysis, significant amount of DNAs from non-
2 target organisms as background exists, and it could affect SNR threshold determination. To
3 examine the effect of such background DNA on the SSR, SBR and SSSDR thresholds, 500 ng of
4 *S. oneidensis* gDNA, or 10 pg per oligonucleotide target was mixed with 1.0 μ g of the yeast
5 gDNA, and their thresholds were determined as described in Fig. 2. With the yeast gDNA as
6 background, the thresholds of SSR and SSSDR for *S. oneidensis* gDNA were determined to be
7 1.75 and 0.65, respectively, which were slightly lower than those without the yeast gDNA as
8 background (Fig. 5A). Similarly, the thresholds of SSR and SSSDR changed from 2.5 and 0.80 to
9 2.0 and 0.70, respectively when synthesized oligonucleotide targets were spiked into the yeast
10 gDNA (Fig. 5B). However, the thresholds of SBR did not change with the target type, or the
11 background DNA (Fig. 5). The results indicate that the thresholds of SSR and SSSDR decreased
12 with the addition of yeast gDNA as background, but that the threshold of SBR stayed the same.

13 To further understand why background DNA caused a decrease in the thresholds of SSR
14 and SSSDR, the changes in signal mean, background mean, and their standard deviations for each
15 spot with the yeast DNA as non-target DNAs were compared with those without the yeast DNA
16 (Fig. 6). When the yeast gDNA was added into the *S. oneidensis* gDNA, the trends of the signal
17 mean and the background mean did not change, but the average signal and background standard
18 deviations increased to 124% and 134%, respectively compared to *S. oneidensis* gDNA only
19 (Fig. 6A). Similarly, when the oligonucleotide targets was used as target templates with the
20 background yeast gDNA, the average signal mean and the average background mean did not
21 change significantly, but both average signal and background standard deviations increased to
22 129% and 148%, respectively in comparison with the oligonucleotide targets only (Fig. 6B). The

1 results indicated that an increase in both signal and background standard deviations might result
2 in lower thresholds of SSR and SDR when non-target DNAs are present.

3 **Determination of SNR thresholds for artificial bacterial mixtures**

4 To examine how DNA mixtures with different compositions affect the SNR threshold
5 determination, *S. oneidensis* gDNA was mixed with other four bacteria in the ratios of (A)
6 **10:1:1:1:1**, (B) **1:1:1:1:1**, and (C) **1:10:10:10:10**, and each mixture had 2.50 µg of gDNA in total.
7 The optimal thresholds of SSR, SBR, and SDR were determined to be 2.00, 1.60, and 0.70,
8 respectively for Mixture (A), and 1.75, 1.60, and 0.60, respectively for Mixture (B) (Table 2).
9 There were only about 23.3% of the defined positive spots were detected on the array for
10 Mixture (C), so no thresholds of SSR, SBR, or SDR could be estimated (Table 2). The results
11 showed that the thresholds of SSR and SDR were decreased with a decrease in the percentage
12 of the target (*S. oneidensis* gDNA) in the sample, but that the thresholds of SBR were not
13 affected, which is also consistent with the results observed with different types of target or with
14 the yeast DNA. It is possible that a decrease in target concentration in a mixed sample may lead
15 to a higher rate for FN or/and FN + FP.

16 **Verification of identified positive spots**

17 To further understand if the identified positive spots based on the above thresholds have
18 significantly higher signals than their backgrounds, the Student t test was used to determine if a
19 probe with replicate spots was positive at a given p value. Since genomic DNA is most
20 commonly used target, this experiment was carried out with *S. oneidensis* MR-1 gDNA (500 ng).
21 The predefined positives (at a low stringency), the t-test identified positives (at p<0.01), and
22 SNR threshold-identified (2.0 for SSR, 1.6 for SBR and 0.70 for SDR) positives were
23 compared, and relatively consistent results were observed (Table 3). Among 368 probes, 60, 249

1 and 59 were defined as positive, negative, and ignored, respectively under a low stringency.
2 Based on t-test, a total of 76 probes were identified as positives with 58 from the defined
3 positives, 3 from the defined negatives, and 15 from the ignored pool at $p < 0.01$ (Table 3).
4 Similar numbers of positives to the t-test analysis were identified based on the SNR thresholds
5 determined above. For example, at the SDR threshold of 0.70, 81, 79, and 75 positives were
6 identified at positive rates of >50%, >70%, and >90%, respectively (Table 3). These results
7 demonstrated that the positive spots or probes identified by SNR thresholds and by the Student t-
8 test were very similar, which was also consistent with the predefined positives and negatives.

9 **Determination of positive spots by SDR threshold for pure culture and soil samples**

10 To demonstrate the application of SDR thresholds for determining positive spots, two sets of
11 data were used. One was pure cultures of wild type (WT) and Δfur mutant (JW707) *Desulfovibrio*
12 *vulgaris* Hildenborough (DvH) with the DvH oligonucleotide microarray (3), and the other was a
13 BioCON soil sample with GeoChip (13). For the first data set, the SDR threshold of 0.80 was
14 used. The average SDR for the *fur* probe was 0.25 for the Δfur mutant, and 2.16 for WT,
15 confirming the absence of this gene in the mutant (Table 4). Fur is a transcriptional regulator,
16 and negatively regulates several genes in the *fur* regulon when it binds to a promoter. The
17 microarray data did show that genes such as *feoA*, *feoB*, *fld*, and *gdp* predicted in the *fur* regulon
18 (25) were up-regulated in the mutant JW707 (Table 4). The Fur regulator has been showed to be
19 involved in oxidative stress responses, which are mainly controlled by the PerR regulator (25).
20 Indeed, our results also showed that *ahpC*, *rbr* and *perR* were over-expressed in the JW707
21 mutant (Table 4). In addition, it was observed that the expression of genes (*cobI*, *COG-fepB*,
22 *fehC*, and *COG-fepD*) involved in iron uptake was repressed, and that the expression of genes
23 (*bfr* and *ftn*) involved in iron storage was induced (Table 4). This is consistent with the fact that

1 more iron may accumulate in the mutant due to the absence of Fur protein. It is noted that
2 different cutoffs for up-regulation and down-regulation were used in this study (two-fold) and
3 the previous study (3).

4 Despite our successful demonstration of SSDR application for pure cultures, a similar
5 demonstration with environmental samples, such as soil, is much more difficult. Thus in this
6 study, we used one soil sample with three hybridizations to see the number of detected positive
7 spots, and their unique and overlap spots among replicates (Table 5). With the thresholds of 2.0
8 for SSR, 1.6 for SBR, and 0.80 for SSDR, the average detected spots were 3858, 4372, and 3828
9 for SSR, SBR, and SSDR, respectively (Table 5). Although the fewest positive spots (3903) were
10 detected by SSDR, it had the highest overlap spot number (3761) and rate (96.3%), but the
11 lowest unique spot number (97) and rate (2.5%), indicating that SSDR is a more accurate method
12 to discriminate true signals from background noise (Table 5). Therefore, the above results
13 demonstrated that the SSDR method with an appropriate threshold could be used to determine
14 positive spots for both pure culture and environmental (e.g. soil) samples.

15

16 **DISCUSSION**

17 How to distinguish a real signal from its background remains challenging in microarray data
18 analysis, and this study focuses on the experimental determination of SNR thresholds. The
19 determination of SNR thresholds is an important step for the generation of high quality
20 microarray data, and its accuracy is critical for the subsequent data processing and biological
21 interpretation of microarray results. Thus this study experimentally determined the thresholds of
22 SNR under different scenarios. The results of this study should provide guidance for users to
23 select appropriate SNR thresholds for their experiments.

1 Considering the standard deviations of pixel intensity of both signal and background, a
2 new calculation method was developed. It has a couple of advantages. First, the signal standard
3 deviation was considered as a parameter together with background standard deviation. Since the
4 pixel intensities of a spot are not uniform, its standard deviation significantly affects the ability
5 of distinguishing a true signal from its background. In this case, a consideration of signal
6 standard deviation can more accurately reflect microarray hybridization behaviors, and more
7 reliably identify a true spot and its threshold. Second, our experimental data demonstrated that
8 fewer false positives and negatives were observed with this method compared to two other
9 methods. SBR did not change with target types, or background DNA since this calculation does
10 not consider signal standard deviation or background standard deviation, but it generally had a
11 high percentage of FN and FP spots, and it may not be a good parameter to distinguish a true
12 signal from its background noise. Therefore, this new method may be used for a general SNR
13 calculation, and more accurate thresholds could be obtained with this calculation.

14 Three possible scenarios, minimizing false positives, minimizing false negatives and
15 optimizing false positives and false negatives, were considered to determine the ranges of SNR
16 thresholds for detecting real signals, but the threshold values for optimal false positives and
17 negatives could be used more often. By optimizing the percentage of FP and FN spots, those
18 thresholds of SSR and SBR determined in this experiment appeared to be lower compared to
19 other commonly accepted thresholds. For example, the threshold of SSR was set to be 3.0 (30),
20 and SBR to be 1.50 (26) or 2.0 (19). Considering all three methods for SNR determination, the
21 ranges of SNR thresholds for gDNA targets were summarized in Table 6. For example, the
22 thresholds of SSR were in the range of 0.5 (no FN), 2.0 (optimal), to 4.0 (no FP), and those of
23 SSDR were in a range of 0.3 (no FN), 0.7 (optimal), to 0.9 (no FP) under a low stringent

1 condition. Those ranges provide a general guideline for users to select appropriate SNR
2 thresholds based on their experiments. There are two points needed to be mentioned. One is that
3 the error rate of 5% (FP + FN) was used in his study, which is considered reasonable since
4 microarray data have relatively high variations due to various reasons, such as the small size,
5 degrees of uniformity of printing pins, and uneven hybridization. The other is that those SNR
6 threshold values determined here for DNA microarray studies under different stringencies and
7 different target types or/and concentrations may only be applied to long (50-70mer)
8 oligonucleotide microarrays. An application of such parameters to short (18-25mer)
9 oligonucleotide microarrays reminds unclear, which needs to be further evaluated.

10 It is known that probe specificity and the stringency of hybridization conditions affect the
11 determination of SNR thresholds. Two stringency conditions were used in this study. As
12 expected, a lower threshold (e.g. SSR = 2.0, SBR = 1.6, and SDR = 0.80) can be used for
13 detecting specific hybridizations under high stringent hybridization conditions (e.g. at a high
14 temperature of 50°C), and a higher threshold (e.g. SSR = 3.0, SBR = 2.0, and SDR = 0.90) may
15 be required for detecting specific hybridizations under low stringent hybridization conditions
16 (e.g. at a low temperature of 42°C).

17 Many factors, such as target type, background DNAs, target composition, and target
18 amount in the tested sample affect the SNR threshold determination. Microarray hybridization
19 signal intensity is determined by the number of probe molecules bound to microarray surface, the
20 number of labeled targets present in the sample, and their ratios, which are closely related to
21 target type and their concentrations. In this study, the synthesized oligonucleotides and PCR
22 amplicons are the simplest target, and both are similar, and they had almost the same thresholds.
23 *S. oneidensis* MR-1 gDNA is more complex, and its threshold was a bit lower. Similarly, the

1 complexity of target is expected to increase in the presence of background DNA, and hence a
2 lower threshold was observed. Further analysis revealed that this might be due to an increase in
3 background standard deviation. This is validated by the fact that the thresholds of SBR did not
4 change with the target type or with background DNA. With the mixed templates, Mixture (A)
5 contained > 70% of real target (*S. oneidensis* gDNA), the threshold did not change significantly.
6 However, a slightly decrease in threshold was observed in Mixture (B) with 20% of real target,
7 and it became undeterminable for Mixture C containing about 2.5% of real target. The the
8 decrease of the thresholds with a decrease of the target template composition can be explained by
9 an increase in sample noise when the target concentration decreased. Sample noise is mostly
10 from labeled molecules in a sample. For example, labeled target solutions can react in a non-
11 specific manner on microarrays, which masks the interactions between a probe and its target and
12 obscures the microarray signal. Therefore, an increase in non-target concentrations leads to an
13 increase in noise, which may reduce SNR thresholds to compromise microarray detectivity. This
14 is also consistent with our observations for different types of target or with background DNA
15 since labeled non-targets such as background DNAs cause a significant amount of background
16 noise.

17 As previous studies showed, the detection limits for 50mer oligonucleotide and 70mer
18 oligonucleotide arrays were estimated to be 25 to 100 ng of gDNA (11) for a pure culture
19 although a higher sensitivity (5~10 ng gDNA) was also observed (24, 29). In the presence of
20 background DNA, the detection limit for 50mer oligonucleotide was estimated to be 50~100 ng
21 of gDNA (24, 29). In the Mixture C, the real target was about 63 ng of gDNA, so it was not
22 surprising that only 23.3% of defined positive probes had true signals. These results suggest that

1 a threshold might change with target compositions, which are closely related to the microarray
2 sensitivity.

3 It is also noted that the amount of target may affect the threshold determination. For
4 example, a higher threshold may be required when a relatively large amount of target is used. In
5 this study, we used the optimal concentrations of 10 pg for each oligonucleotide, 100 pg for each
6 PCR amplicon, and 500 ng for gDNA, which are considered equivalent amounts of the target in
7 samples. This is a simulation for a pure culture, or a mixture of a few known microorganisms.
8 For a sample with many unknown microorganisms, such as microbial communities in soil and
9 the human intestinal tract, a determination of SNR thresholds may be even more challenging.
10 Because of unequal abundance, low-abundant genes/microorganisms may not be detected even at
11 a relatively low threshold.

12 In summary, three methods were used to calculate SNR values, and the newly developed
13 calculation showed a better performance for distinguishing a true signal from its background
14 than the other two methods. The positives identified based on SNR thresholds were verified by
15 the Student t-test across many replicate data, and consistent results were obtained. This study
16 provides guidance for the selection of SNR thresholds for different samples, such as PCR
17 amplicons, and genomic DNA from pure cultures and simple mixed cultures.

18

19 **ACKNOWLEDGEMENTS**

20 The authors thank Meiyang Xu for providing GeoChip data of the BioCON soil sample and
21 Yuting Liang for providing *Desulfovibrio vulgaris* Hildenborough microarray data of both wild
22 type and the *Δfur* mutant. This research was supported by The United States Department of

- 1 Energy under Genomics:GTL program through the Virtual Institute of Microbial Stress and
- 2 Survival (VIMSS; <http://vimss.jbl.gov>), and the Environmental Remediation Science Program.

1 **REFERENCES**

- 2 1. **Aakra, A., O. L. Nyquist, L. Snipen, T. S. Reiersen, and I. F. Nes.** 2007. Survey of
3 genomic diversity among *Enterococcus faecalis* strains by microarray-based comparative
4 genomic hybridization. *Appl. Environ. Microbiol.* **73**: 2207-2217.
- 5 2. **Basarsky, T., D. Verdnik, D. Willis, and J. Zhai.** 2000. An overview of a DNA microarray
6 scanner: design essentials for an integrated acquisition and analysis platform. In M. Skena
7 (ed.), *Microarray Biochip Technology*, BioTechniques Books, Eaton Publishing.
- 8 3. **Bender, K.S., B.C.B Yen, C.L. Hemme, Z. Yang, Z. He, Q. He, J. Zhou, K.H. Huang,**
9 **E.J. Alm, T.C. Hazen, A.P. Arkin, and J.D. Wall.** 2007. Analysis of a ferric uptake
10 regulator (Fur) mutant of *Desulfovibrio vulgaris* Hildenborough. *Appl. Environ. Microbiol.*
11 **73**: 5389-5400.
- 12 4. **Bodrossy, L., and A. Sessitsch.** 2004. Oligonucleotide microarrays in microbial diagnostics.
13 *Curr. Opin. Microbiol.* **7**:245–254.
- 14 5. **Bodrossy, L., N. Stralis-Pavese, M. Konrad-Köszler, A. Weilharter, T. G. Reichenauer,**
15 **D. Schöfer, and A. Sessitsch.** 2006. mRNA-based parallel detection of active methanotroph
16 populations by use of a diagnostic microarray. *Appl. Environ. Microbiol.*, **72**: 1672-1676.
- 17 6. **Brodie, E. L., T. Z. DeSantis, J. P. M. Parker, I. X. Zubietta, Y. M. iceno, and G. L.**
18 **Andersen.** 2007. Urban aerosols harbor diverse and dynamic bacterial populations. *Proc.*
19 *Natl. Acad. Sci. USA* **104**: 299-304.
- 20 7. **Debouck, C., and P. N. Goodfellow.** 1999. DNA microarrays in drug discovery and
21 development. *Nat. Genet.* **21**: 48-50.

- 1 8. **Dziejman, M., E. Balon, D. Boyd, C. M. Fraser, J. F. Heidelberg, and J. J. Mekalanos.**
2 2002. Comparative genomic analysis of *Vibrio cholerae*: genes that correlate with cholera
3 endemic and pandemic disease. *Proc. Natl. Acad. Sci. USA* **99**: 1556–1561.
- 4 9. **Gresham, D., D. M. Ruderfer, S.C. Pratt, J. Schacherer, M. J. Dunham, D. Botstein,**
5 **and L. Kruglyak.** 2006. Genome-Wide Detection of Polymorphisms at Nucleotide
6 Resolution with a Single DNA Microarray. *Science* **311**: 1932-1936.
- 7 10. **Han, W., B. Liu, B. Cao, L. Beutin, U. Krüger, H. Liu, Y. Li, Y. Liu, L. Feng, and L.**
8 **Wang.** 2007. DNA Microarray-Based Identification of Serogroups and Virulence Gene
9 Patterns of *Escherichia coli* Isolates Associated with Porcine Postweaning Diarrhea and
10 Edema Disease. *Appl. Environ. Microbiol.* **73**: 4082-4088.
- 11 11. **He, Z., L. Wu, M. W. Fields, and J. Zhou.** 2005a. Comparison of microarrays with
12 different probe sizes for monitoring gene expression. *Appl. Environ. Microbiol.* **71**: 5154-
13 5162.
- 14 12. **He, Z., L. Wu, X. Li, M. W. Fields, and J. Zhou.** 2005b. Empirical establishment of
15 oligonucleotide probe design criteria. *Appl. Environ. Microbiol.* **71**: 3753-3760.
- 16 13. **He, Z., T. J. Gentry, C. W. Schadt, L. Wu, J. Liebich, S. C. Chong, Z. Huang, W. Wu,**
17 **B. Gu, P. Jardine, C. Criddle, and J. Zhou.** 2007. GeoChip: A comprehensive microarray
18 for investigating biogeochemical, ecological, and environmental processes. *ISME J* **1**: 67-77.
- 19 14. **Hinchliffe, S. J., K. E. Isherwood, R. A. Stabler, M. B. Prentice, A. Rakin, R. A. Nichols,**
20 **P. C. Oyston, J. Hinds, R. W. Titball, and B. W. Wren.** 2003. Application of DNA
21 microarrays to study the evolutionary genomics of *Yersinia pestis* and *Yersinia*
22 *pseudotuberculosis*. *Genome Res.* **13**: 2018–2029.

- 1 15. **Hoffman, C.S., and S. Winston.** 1987. A ten-minute DNA preparation from yeast
2 efficiently releases autonomous plasmids for transformiaon of *Escherichia coli*. *Gene* **57**:
3 267-272.
- 4 16. **Kim, I. J., H. C. Kang, S. G. Jang, S. A. Ahn, H. J. Yoon, and J. G. Park.** 2007.
5 Development and applications of a BRAF oligonucleotide microarray. *J. Mol. Diagn.* **9**: 55-
6 63.
- 7 17. **Liebich, J., C.W. Schadt, S.C. Chong, Z. He, S.K. Rhee, and J. Zhou.** 2006.
8 Improvement of oligonucleotide design criteria for the development of functional gene
9 microarrays for environmental applications. *Appl. Environ. Microbiol.* **72**:1688-1691.
- 10 18. **Loy, A., A. Lehner, N. Lee, J. Adamczyk, H. Meier, J. Ernst, K. Schleifer, and M.**
11 **Wagner.** 2002. Oligonucleotide microarray for 16S rRNA gene-based detection of all
12 recognized lineages of sulfate-reducing prokaryotes in the environment. *Appl. Environ.*
13 *Microbiol.* **68**: 5064-5081.
- 14 19. **Loy, A., C. Schulz, S. Lücker, A. Schöpfer-Wendels, K. Stoecker, C. Baranyi, A.**
15 **Lehner, and M. Wangner.** 2005. 16S rRNA gene-based oligonucleotide microarray for
16 environmental monitoring of the betaproteobacterial order “*Rhodocyclales*”. *Appl. Environ.*
17 *Microbiol.* **71**: 1373-1386.
- 18 20. **Maynard, C., F. Berthiaume, K. Lemarchand, J. Harel, P. Payment, P. Bayardelle, L.**
19 **Masson, and R. Brousseau.** 2005. Waterborne pathogen detection by use of
20 oligonucleotide-based microarrays. *Appl. Environ. Microbiol.* **71**: 8548–8557.

- 1 21. **Quiñones, B., Parker, C.T., Janda, Jr., J.M., Miller, W.G., and Mandrell, R.E.** 2007.
2 Detection and genotyping of *Arcobacter* and *Campylobacter* isolates from retail chicken
3 samples by use of DNA oligonucleotide arrays. *Appl. Envir. Microbiol.* **73**: 3645-3655.
- 4 22. **Ragoussis, J., and G. Elvidge.** 2006. Affymetrix GeneChip System: Moving from Research
5 to the Clinic. *Expert Rev. Mol. Diagn.* **6**: 145-52.
- 6 23. **Reich, P. B., J. Knops, D. Tilman, J. Craine, D. Ellsworth, M. Tjoelker, T. Lee, D.**
7 **Wedin, S. Naeem, D. Bahauddin, G. Hendrey, S. Jose, K. Wrage, J. Goth, and W.**
8 **Bengston.** 2001. Plant diversity enhances ecosystem responses to elevated CO₂ and nitrogen
9 deposition. *Nature* **410**: 809-812.
- 10 24. **Rhee, S.K., X. Liu, L. Wu, S. C. Chong, X. Wan, and J. Zhou.** 2004. Detection of
11 biodegradation and biotransformation genes in microbial communities using 50-mer
12 oligonucleotide microarrays. *Appl. Environ. Microbiol.* **70**: 4303-4317.
- 13 25. **Rodionov, D. A., I. Dubchak, A. P. Arkin, E. Alm, and M. S. Gelfand.** 2004.
14 Reconstruction of regulatory and metabolic pathways in metal-reducing δ -proteobacteria.
15 *Genome Biol.* **5**: R90.
- 16 26. **Schena, M.** 2003. *Microarray Analysis*, John Wiley's & Sons, Inc., Hoboken, NJ.
- 17 27. **Small, J., D. R. Call, F. J. Brockman, T. M. Straub, and D. P. Chandler.** 2001. Direct
18 detection of 16S rRNA in soil extracts by using oligonucleotide microarrays. *Appl. Environ.*
19 *Microbiol.* **67**: 4708-4716.
- 20 28. **Taroncher-Oldedburg, G., E. M. Griner, C. A. Francis, and B. B. Ward.** 2003.
21 Oligonucleotide microarray for the study of functional gene diversity in the Nitrogen cycle in
22 the environment. *Appl. Environ. Microbiol.* **69**: 1159-1171.

- 1 29. **Tiquia, S.M., L. Wu, S. C. Chong, S. Passovets, D. Xu, Y. Xu, and J. Zhou.** 2004.
2 Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in
3 environmental samples. *Biotechniques* **36**: 664-675.
- 4 30. **Verdick, D., S. Handran, and S. Pickett.** 2002. Key considerations for accurate microarray
5 scanning and image analysis. In G. Kamberova (ed.), *DNA Array Image Analysis: Nuts and*
6 *Bolts*, DNA Press LLC, Salem, MA. Pp 83-98.
- 7 31. **Vora, G. J., C. E. Meador, M. M. Bird, C. A. Bopp, J. D. Andreadis, and D. A. Stenger.**
8 2005. Microarray-based detection of genetic heterogeneity, antimicrobial resistance, and the
9 viable but nonculturable state in human pathogenic *Vibrio spp.* *Proc. Natl. Acad. Sci. USA*
10 **102**: 19109-19114.
- 11 32. **Wu, L., D. K. Thompson, X. D. Liu, M. W. Fields, C. E. Bagwell, J. M. Tiedje, J. Zhou.**
12 2004. Development and evaluation of microarray-based whole-genome hybridization for
13 detection of microorganisms within the context of environmental applications. *Environ. Sci.*
14 *Technol.* **38**: 6775-6782.
- 15 33. **Zhou, J.** 2003. Microarrays for bacterial detection and microbial community analysis. *Curr.*
16 *Opin. Microbiol.* **6**: 288-294.
- 17 34. **Zhou, J., M. R. Fries, J. C. Chee-Sanford, and J. M. Tiedje.** 1995. Phylogenetic analyses
18 of a new group of denitrifiers capable of anaerobic growth on toluene: Description of
19 *Azoarcus tolulyticus* sp. nov. *Int. J. Syst. Bacteriol.* **45**: 500-506.
- 20 35. **Zhou, S., K. Kassauei, D. J. Cutler, G. C. Kennedy, D. Sidransky, A. Maitra, and J.**
21 **Califano.** 2006. An oligonucleotide microarray for high-throughput sequencing of the
22 mitochondrial genome. *J. Mol. Diagn.* **8**: 476-482.

1 **FIGURE LEGENDS**

2

3 **Fig. 1.** Schematic presentation of the SSDR calculation method. A, B and C represent
4 SSDR<1.0, SSDR = 1.0 and SSDR >1.0, respectively. All the four parameters used in
5 calculation extracted from the ImaGene output files ([Manual of ImaGene](#)).

6

7 **Fig. 2.** Determination of thresholds of SSR (A), SBR (B), and SSDR (C) at a low stringency by
8 minimizing the percentage of false positive and false negative spots. 10 pg of each synthesized
9 oligonucleotide was used to hybridize with the array, and 5 replicate slides were used. SSR, SBR
10 and SSDR were determined to be 2.5, 1.6, and 0.80, respectively.

11

12 **Fig. 3.** Determination of thresholds of SSR (A), SBR (B), and SSDR (C) at a high stringency by
13 minimizing the percentage of false positive and false negative spots. 10 pg of each synthesized
14 oligonucleotide was used to hybridize with the array, and 5 replicate slides were used. SSR, SBR
15 and SSDR were determined to be 3.0, 2.0, and 0.90, respectively.

16

17 **Fig. 4.** Effects of target types on the thresholds and the percentages of false positive (FP), false
18 negatives (FN), and both (FP+FN) for SSR (A), SBR (B), and SSDR (C). The left y-axis
19 presents the optimal threshold, and the right y-axis presents the percentage of FP, FN, or FP +
20 FN under the optimal threshold. Targets used were synthesized oligonucleotides (10 pg each),
21 PCR amplicons (100 pg each), and *S. oneidensis* MR1 gDNA (500 ng). The more significant p
22 value is shown on the top of each column with the following notions: nd=no difference,

1 *= $p < 0.10$, **= $p < 0.05$, and ***= $p < 0.01$ (the Student t test) when one type of target was
2 compared with two others.

3
4 **Fig. 5.** Effects of background DNA on the determination of SSR, SBR and SDR thresholds.
5 500 ng of *S. oneidensis* MR-1 gDNA (A) and 10 pg for each synthesized oligonucleotide (B)
6 were spiked into 1.0 μ g of yeast gDNA. For synthesized oligonucleotide targets, the yeast gDNA
7 was first labeled and then mixed with the spiked oligonucleotides. *S. oneidensis* MR-1 gDNA
8 was first mixed with the yeast gDNA, and then labeled together. The significance is shown on
9 the top of each column with the following notions: nd=no difference, *= $p < 0.10$, **= $p < 0.05$, and
10 ***= $p < 0.01$ (the Student t test) when thresholds with background DNA were compared to those
11 without background DNA.

12
13 **Fig. 6.** Comparison of changes in signal mean, background mean, signal standard deviation and
14 background standard deviation for each spot on the array when the yeast gDNA was added into
15 the *S. oneidensis* gDNA (A), or the synthesized oligonucleotide targets (B).

16

1 **Table 1.** The thresholds of SSR, SBR and SSDR determined by minimizing the percentage of
 2 false positive (FP) spots or false negative (FN) spots on the array using synthesized
 3 oligonucleotide targets under low and high stringencies.
 4

A. The thresholds of SSR, SBR and SSDR at the defined low stringency			
	SSR	SBR	SSDR
No FP	5.0	5.0	1.00
1% FP	4.0	3.5	0.90
5% FP	2.0	1.8	0.70
5% FN	2.5	1.6	0.85
1% FN	1.5	1.2	0.70
No FN	0.5	0.5	0.30
B. The thresholds of SSR, SBR and SSDR at the defined high stringency			
	SSR	SBR	SSDR
No FP	5.0	5.0	1.10
1% FP	4.5	4.0	1.00
5% FP	2.5	2.0	0.70
5% FN	3.0	1.8	0.95
1% FN	2.0	1.4	0.75
No FN	1.0	1.0	0.50

5

1 **Table 2.** The thresholds of SSR, SBR and SSDR and the percentages of false positives, false
 2 negatives, or both for artificial bacterial mixtures. Genomic DNAs from Mixture A, B and C
 3 containing *S. oneidensis* MR-1 (bold)) and other four bacteria with different ratios were used as
 4 targets. SSR, SBR, SSDR, and percentages of false positives and false negatives were
 5 determined as described in Fig. 2. Five slides were used.
 6

		Mixture A (10 :1:1:1:1)	Mixture B (1 :1:1:1:1)	Mixture C (1 :10:10:10:10)
No. of defined positive spots		300	300	300
% of detected positive spots		318	311	70
SSR	Threshold	2.0	1.75	ND
	% of false positives (FP)	4.3	3.5	0
	% of false negatives (FN)	3.3	3.4	76.7
	% of total FP and FN	7.6	6.9	76.7
SBR	Threshold	1.60	1.60	ND
	% of false positives (FP)	4.7	3.6	0
	% of false negatives (FN)	3.3	4.7	76.7
	% of total FP and FN	8.0	8.3	76.7
SSDR	Threshold	0.70	0.60	ND
	% of false positives (FP)	2.7	2.2	0
	% of false negatives (FN)	2.8	3.7	76.7
	% of total FP and FN	5.5	5.9	76.7

7

1 **Table 3.** Comparison of positive probes identified by probe design criteria, by the Student t-test,
 2 and by SNR thresholds. 368 probes were valid for analysis when 500 ng of labeled *S. oneidensis*
 3 MR1 gDNA hybridized with the array. Five slides were used with four replicates in each slide, so
 4 each probe had up to 20 spots.
 5

A. Defined and t-test identified positive probes at the low stringency ($p < 0.01$ for the Student t test)				
	No. of probes			
Defined positives	60 ^a			
Defined negatives	249 ^b			
Ignored	59 ^c			
No. of t-test positives	57 ^a + 4 ^b + 15 ^c = 76			
No. of t-test negatives	3 ^a + 245 ^b + 44 ^c = 292			
B. SNR-threshold-identified positive probes at different positive rates (PR = the number of positive spots identified by SNR thresholds *100/total number of spots for each probe).				
	Threshold	PR > 50%	PR > 70%	PR > 90%
No. of SSR-identified positives (% of t-test positives)	2.0	58 ^a +7 ^b +21 ^c =86 (113%)	58 ^a +5 ^b +19 ^c =82 (108%)	57 ^a +3 ^b +18 ^c =78 (103%)
No. of SBR-identified positives (% of t-test positives)	1.6	58 ^a +8 ^b +25 ^c =91 (120%)	57 ^a +6 ^b +23 ^c =86 (113%)	56 ^a +4 ^b +20 ^c =80 (105%)
No. of SSDR-identified positives (% of t-test positives)	0.70	59 ^a +4 ^b +18 ^c =81 (107%)	59 ^a +3 ^b +17 ^c =79 (104%)	58 ^a +1 ^b +16 ^c =75 (99%)

6
 7 ^athe size of defined positive probe pool; ^bthe size of defined negative probe pool; ^cthe number of
 8 ignored probes based on defined positive, negative, and ignored probe pools.
 9

1 **Table 4.** Examples of transcriptional changes of function-known genes in *Δfur* mutant (JW707)
 2 and wild type (WT) of *D. vulgaris* Hildenborough.
 3

Category/ DVU	Gene	Annotated function	SSDR ^a		Expression ratio (JW707/WT)
			JW707	WT	
Genes in the predicted Fur regulon^b					
DVU0303	<i>genZ</i>	GenZ, hypothetical protein	2.16±0.285 ^c	1.78±0.172 ^c	2.11
DVU0304	<i>genY</i>	GenY, hypothetical protein	2.47±0.277	2.15±0.122	2.27
DVU0763	<i>gdp</i>	GGDEF domain protein	2.28±0.321	1.96±0.231	4.08
DVU0942	<i>fur</i>	Fur, transcriptional regulator	0.25±0.036	2.16±0.116	ND ^d
DVU2571	<i>feoB</i>	Ferrous iron transport protein B	1.97±0.166	1.95±0.142	1.96
DVU2572	<i>feoA</i>	Ferrous iron transport protein A	1.72±0.321	1.83±0.211	1.76
DVU2574	<i>feoA</i>	Ferrous ion transport protein	2.17±0.277	1.93±0.102	2.67
DVU2680	<i>fld</i>	Flavodoxin	1.88±0.130	2.06±0.133	1.50
Genes in the predicted PerR regulon^b					
DVU2247	<i>ahpC</i>	Antioxidant, AhpC/Tsa family	2.02±0.220	2.03±0.186	2.12
DVU2318	<i>rbr</i>	Rubrerhythrin, putative	2.47±0.277	1.76±0.122	2.94
DVU3095	<i>perR</i>	PerR, transcriptional regulator	1.88±0.213	1.90±0.133	1.61
Other iron-related genes					
DVU0646	<i>cobI</i>	Precorrin-2 C20-methyltransferase	1.28±0.096	2.35±0.182	0.30
DVU0647	<i>COG-fepB</i>	Iron compound ABC transporter, iron-binding protein	0.93±0.071	2.11±0.171	0.14
DVU0648	<i>fepC</i>	Iron compound ABC transporter, ATP-binding protein	1.17±0.277	1.84±0.132	0.25
DVU0649	<i>COG-fepD</i>	Iron compound ABC transporter, permease protein	1.20±0.076	2.26±0.119	0.28
DVU1397	<i>bfr</i>	Bacterioferritin	2.27±0.217	2.16±0.212	1.97
DVU1568	<i>fn</i>	Ferritin	1.89±0.173	2.33±0.222	1.89

4
 5 a. SDR was calculated from Cy5-labeled cDNA signal while Cy3-labeled gDNA was used for both
 6 JW707 and WT; b. Those are predicted by Rodionov et al. (25); c. Mean±SD (n=6); d. Not determined
 7 due to the lack of Cy5 signal of the *Δfur* mutant.
 8

1 **Table 5.** The number of detected, unique and overlap spots among replicate A, B, and C. Three
 2 different methods, SSR, SBR and SSDR and their pre-determined thresholds were used for the
 3 detection of positive spots.
 4

	SSR	SBR	SSDR
Threshold	2.0	1.6	0.80
Detected positive spots (mean \pm sd)	3858 \pm 157	4372 \pm 322	3828 \pm 60
Total positive spots ($A \cup B \cup C$)	4132	4743	3903
a. Unique positive spots among three replicates	232 (5.6%)	566 (12%)	97 (2.5%)
b. Overlapped positive spots among two replicates [$(A \cap B) \cup (A \cap C) \cup (B \cap C)$]	263 (6.4%)	521 (11%)	45 (1.2%)
c. Overlapped positive spots among three replicates ($A \cap B \cap C$)	3637 (88%)	3656 (77%)	3761 (96.3%)

5

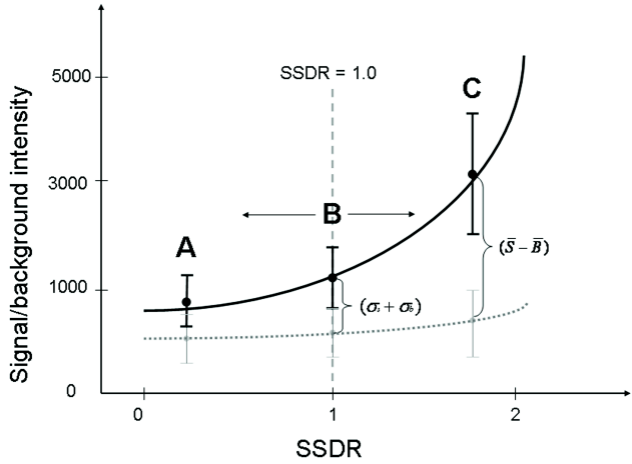
1 **Table 6.** A summary of the ranges of experimentally determined SNR threshold under low and
2 high stringent conditions using the *S. oneidensis* MR1 gDNA target.

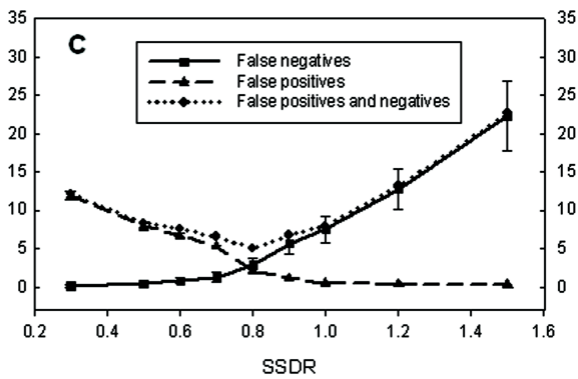
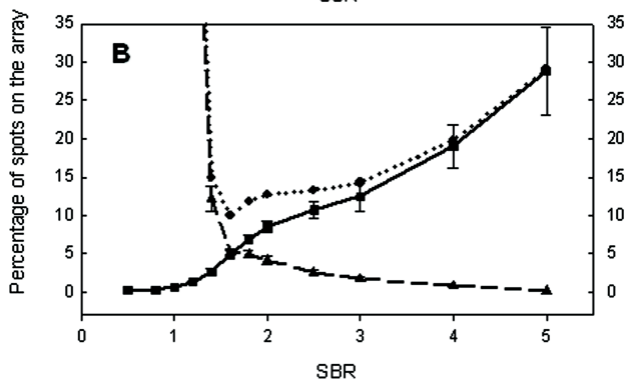
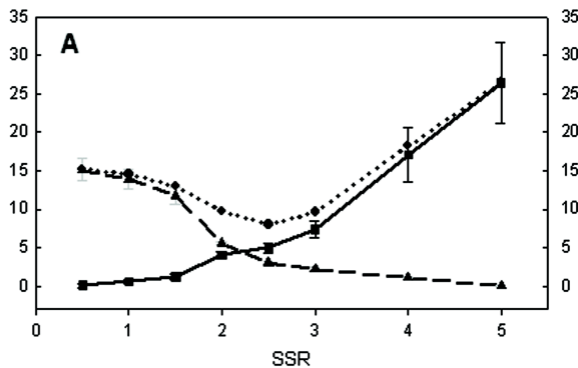
3

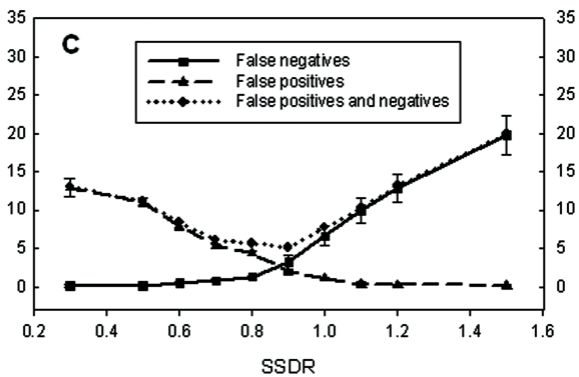
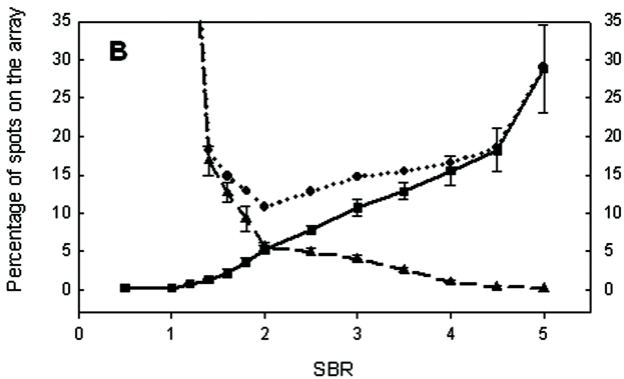
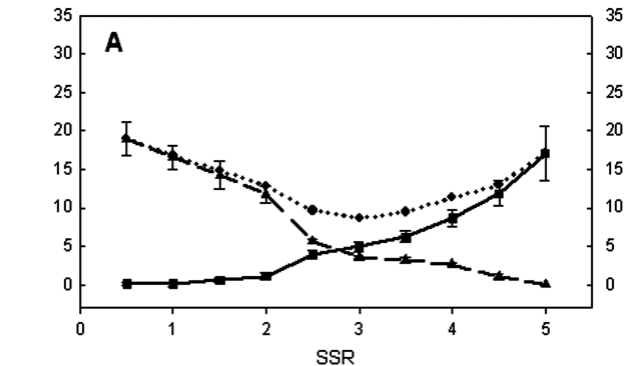
	Low stringency			High stringency		
	No FN	Optimal	No FP	No FN	Optimal	No FP
SSR	0.5	2.0	4.0	1.0	2.5	4.5
SBR	0.5	1.6	4.0	1.1	1.8	4.5
SSDR	0.3	0.7	0.9	0.4	0.8	1.0

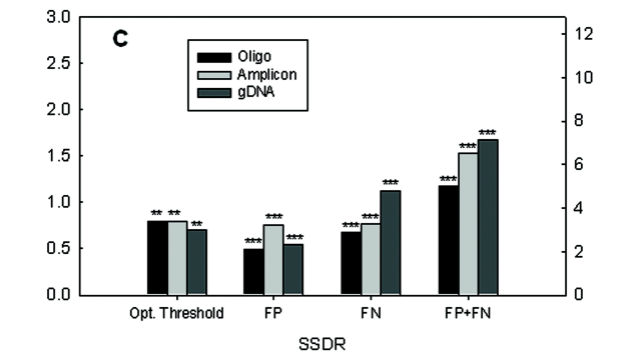
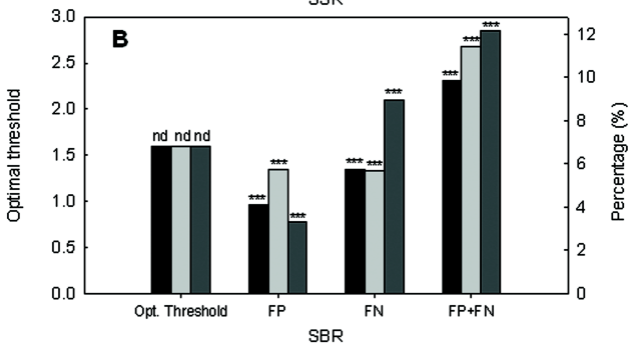
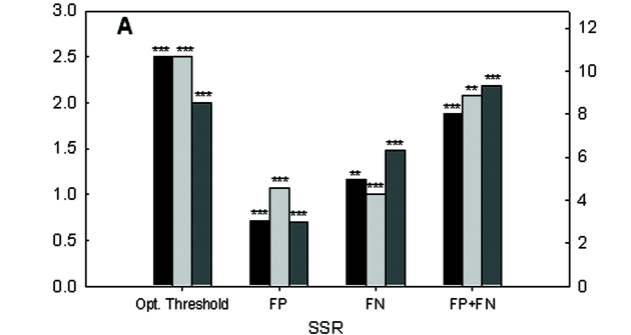
4

5









Threshold

