BMC Genomics

Research article

Improving signal intensities for genes with low-expression on oligonucleotide microarrays

Latha Ramdas, David E Cogdell, Jack Y Jia, Ellen E Taylor, Valerie R Dunmire, Limei Hu, Stanley R Hamilton and Wei Zhang*

Address: Cancer Genomics Core Laboratory, Department of Pathology, The University of Texas M. D. Anderson Cancer Center Houston, TX 77030, USA

Email: Latha Ramdas - lramdas@mdanderson.org; David E Cogdell - dcogdell@mdanderson.org; Jack Y Jia - jiayu@mdanderson.org; Ellen E Taylor - etaylor@mdanderson.org; Valerie R Dunmire - vdunmire@mdanderson.org; Limei Hu - lhu@mdanderson.org; Stanley R Hamilton - shamilton@mdanderson.org; Wei Zhang* - wzhang@mdanderson.org

* Corresponding author

Abstract

Background: DNA microarrays using long oligonucleotide probes are widely used to evaluate gene expression in biological samples. These oligonucleotides are pre-synthesized and sequence-optimized to represent specific genes with minimal cross-hybridization to homologous genes. Probe length and concentration are critical factors for signal sensitivity, particularly when genes with various expression levels are being tested. We evaluated the effects of oligonucleotide probe length and concentration on signal intensity measurements of the expression levels of genes in a target sample.

Results: Selected genes of various expression levels in a single cell line were hybridized to oligonucleotide arrays of four lengths and four concentrations of probes to determine how these critical parameters affected the intensity of the signal representing their expression. We found that oligonucleotides of longer length significantly increased the signals of genes with low-expression in the target. High-expressing gene signals were also boosted but to a lesser degree. Increasing the probe concentration, however, did not linearly increase the signal intensity for either low- or high-expressing genes.

Conclusions: We conclude that the longer the oligonuclotide probe the better the signal intensities of low expressing genes on oligonucleotide arrays.

Background

DNA microarray technology allows analysis of the expression of thousands of genes in a single experiment [1]. Most microarray fabrications spot either the cDNA [polymerase chain reaction (PCR) products] or a long oligonucleotide probe for each gene onto a solid support, such as a chemically coated glass slide. In recent years, long oligonucleotide microarrays have become more popular than cDNA arrays because the generation of cDNA microarrays involves many laborious and error-prone steps, including bacterial culturing, PCR amplification and purification, and DNA sequence validation [2,3]. In addition, cDNA microarrays are prone to cross-hybridization with gene family members that have 70% or more sequence homology [4].

For long oligonucleotide microarrays, pre-synthesized oligonucleotides are used as probes to be spotted onto chemically coated glass slides [5-7]. The oligonucleotides are synthesized by phosphoramidite synthesis in which the oligonucleotides have 99.4% coupling efficiency, which measures how efficiently the DNA synthesizer adds next nucleotide to the growing oligonucleotide. A 99.4 % efficiency indicates that at every coupling step approximately 0.6 % of the available nucleotides fail to react. The percentage of full length oligonucleotide depends on coupling efficiency. Synthesis of a 70-base oligonucleotide yields more than 65% full-length product (% full length = (coupling efficiency)$^{n-1}$, where n = total number of nucleotides). Several sequenced genomes are already available, and more will become available over the next few years. The oligonucleotides can be selected from the least homologous region of the gene as determined by a BLAST search of the human genome sequence [8], thus increasing the specificity of hybridization.

Several factors are considered when determining the optimal length of long oligonucleotides for microarrays. In general, the longer the oligonucleotides are, the more efficient is the hybridization [9]. One study suggests that length-dependent hybridization efficiency reaches a plateau at 712 bases for PCR products, above which the effect of length on hybridization rate decreases [10]. However, current oligonucleotide synthesis technologies have limitations far below this hybridization efficiency limit, and the efficiency of generating full-length oligonucleotide decreases as the length increases [11,12]. As the length approaches 100 bases, a 99.5 % coupling efficiency in synthesis will yield only 61 % of full length product and this drops dramatically to 37 % when the coupling efficiency drops to 99 % [13,14]. This sets a limitation to the synthesis of oligonucleotides. In addition, the cost of oligonucleotide synthesis increases with their length. Commercial libraries contain oligonucleotides of lengths ranging from 50 to 80 bases. Several studies have used oligonucleotides of 50 nucleotides, and some compared the hybridization behaviors of oligonucleotides of 50 and 70 nucleotides [15].

To gain insight into the behavior of long oligonucleotide microarrays, we performed a comparative study evaluating the effects of oligonucleotide length and the amounts of oligonucleotides printed. In this study, we investigated microarray behavior using pre-synthesized, unmodified oligonucleotides deposited on glass slides. We systematically evaluated the effect of lengths of the oligonucleotides signal intensities of genes with different expression levels in the target sample. We also studied how the concentration of the oligonucleotide probes of various lengths affected the signal intensities of these genes.

Results and discussion

In microarray experiment, reliable measurement is more achievable for highly expressed genes in a target sample than for those expressed at low levels. Accurate measurement of low-expressing genes is challenging because the low-intensity signals are not only weaker, but also more variable [5,17].

Several studies have suggested that long oligonucleotides, as long as 50 nucleotides, give satisfactory microarray results [5,15]. However, most of these studies have focussed on high-expressing genes that have high signal intensities on microarrays. This approval may bias the conclusions because low-expressing genes pose more challenge to accurate microarray measurement and analyses. To assess whether low-expressing genes can be tested accurately in microarrays, we included several genes that are expressed at low levels in the target cell line. Thirty genes were selected based on multiple microarray data from RKO colon cancer cell line. The expression levels were categorized as high-expressing genes when the average signal/noise (S/N) ratio were above 50, as medium-expressing genes when the average S/N were between 15 and 50, low-expressing genes when the average S/N were between 2 and 5 and as no-expression genes when the S/N is less than 2. All 30 genes were spotted at four concentrations (20, 30, 40 and 50 μM) and four lengths (30, 40, 50 and 70 nucleotides) onto poly-L-lysine coated glass slides. Thus, a total of 240 oligonucleotide samples were hybridized and analyzed.

Figure 1a is the image obtained after the oligonucleotide probes were hybridized with the Cyanine-5 labeled target cDNA. Each grid represents a single gene, and the spotting is in duplicate as shown in Figure 1b. The grid number, the corresponding gene description, and expected expression level of the genes on the microarray in the target cell line are shown in Table 1.

Oligonucleotide length and expression levels

Figure 2 shows the signal intensities of the spots at each of the four lengths at 50 μM probe concentration. For those genes expressed in the target cell line, the signals from 70-base oligonucleotides were of greater intensity than those obtained using shorter probes. Our data also suggests that this increase was more pronounced for the low-expressing genes. The average absolute increase in signal intensities for genes with high and low-expression levels, as a function of increasing length, the base being 30, is shown in Table 2. Similarly, the average intensity ratios for increasing concentrations of low and high expressing genes with respect to 20 μM are also given in Table 2.

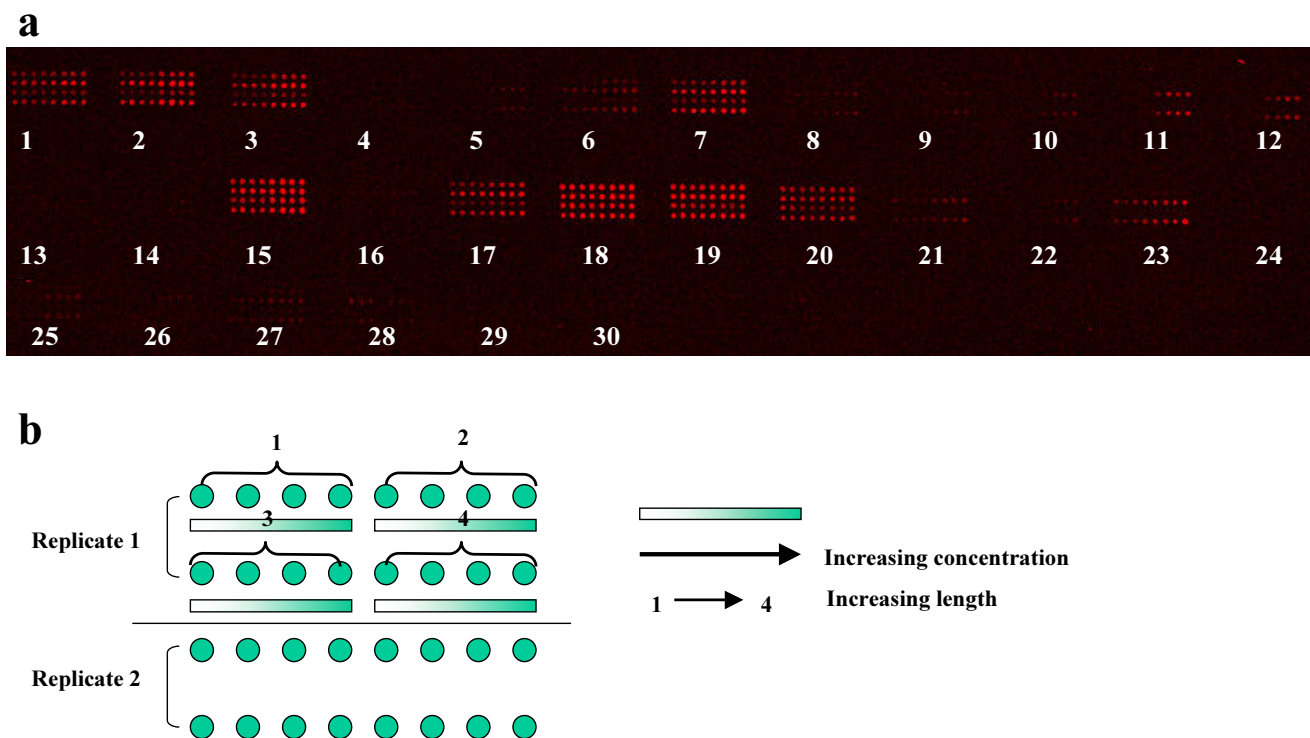Figure 3 shows the relative fold-increase in signal intensity for increasing oligonucleotide length compared to the

Figure 1
(a) Cyanine-5 labeled image of the spotted oligonucleotides hybridized with RNA from RKO cell line. (b) Pattern of spotting of the probes.

shortest oligonucleotides of 30 bases. Evaluation of the effect of oligonucleotide length indicates that there is a close to linear increase in signal intensity as the length of the probe increases (Table 2). For low-expressing genes, the signal intensity yielded by hybridization with a 70-base nucleotide probe is up to 120 times that of the baseline value (Fig. 3a). This increase is much greater than that conferred by 50- and 40-base nucleotides, in concordance with the earlier observation that 70-base probes gave the optimum signal. Similarly, high-expressing genes showed higher signal intensities with the 70-nucleotide probe, although the relative intensity increase was to a maximum of 18 times the baseline value (Fig. 3b). Thus, the long-oligonucleotide probes gave better signal intensities than the shorter probes, and the low-expressing genes had the greatest improvement. The improvement is statistically significant with p-values < 0.05 using the t-test of two samples with unequal variance and a two-tailed test.

Probe concentration and signal intensity

We evaluated the effect of concentration of the printed oligonucleotide probes on signal intensity after hybridization. There are always concerns that the concentration of the probe might influence the signal output for low-

expressing genes, although this is not a major issue for the medium- and high-expressing genes. We reasoned that an increase in the number of probe molecules on the array may help capture the rare target cDNAs and enable their detection by imaging. Unlike the probe length analysis, in which it was assumed that the probe length itself would affect the hybridization, in the concentration studies it was assumed that the probes on the slide would generally be printed in excess concentration to the amount of cDNA in the target sample. Therefore, we expect the signal intensities to show a linear range dependence on the target cDNA concentration that follows a pseudo-first-order kinetics model instead of a second order kinetics model [10,17].

Figure 4 shows the relative increase in signal as a function of probe concentrations. The data from a set of nine low-expressing genes show that for the samples used in this study, the effect of concentration on signal intensity was minimal, with a maximum of only five times the baseline intensity yielded by the highest probe concentration (Fig. 4a). Although significant, this increase is a much smaller change compared to the 120-fold difference observed as length of the probe increases. Similar, but even less

Table 1: Relative Expression levels of selected genes in target RKO colon cancer cell line

Grid #	ID	Gene Description	Expression in target cell line
1	**Co1**	beta actin	+++
2	**Co2**	GAPDH	+++
3	**Co3**	Tubulin	+++
4	**HK1**	Ferritin heavy chain	-
5	**HK10**	Homo sapiens keratin, hair, basic, 5 (KRTHB5) mRNA	+
6	**HK11**	Homo sapiens lysophospholipase II (LYPLA2), mRNA	-
7	**HK12**	Homo sapiens kinesin family protein 3B (KIF3B) mRNA	+++
8	**HK13**	Homo sapiens tubulin-specific chaperone c (TBCC), mRNA	+
9	**HK14**	Homo sapiens tankyrase, TRF1-interacting ankyrin-related ADP-ribose polymerase (TNKS) mRNA, and translated products	-
10	**HK15**	Homo sapiens nesca protein (NESCA), mRNA	+
11	**HK16**	Homo sapiens myosin 5C (MYO5C), mRNA	++
12	**HK17**	Human myosin heavy chain 12 (MYO5A) mRNA, complete cds	+
13	**HK18**	Homo sapiens reticulon 2 (RTN2) mRNA	-
14	**HK19**	Homo sapiens keratin, hair, acidic,2 (KRTHA2), mRNA	-
15	**HK2**	Ferritin L chain	+++
16	**HK20**	Homo sapiens keratin 13 (KRT13) mRNA	+
17	**HK3**	Phospholipase A2	+++
18	**HK4**	Ribosomal Protein S12	+++
19	**HK5**	Ribosomal protein L 12	+++
20	**HK6**	Ribosomal protein S9	++
21	**HK7**	hypoxanthine-guanine phophoribosyltranferase	++
22	**HK8**	H. sapiens mRNA for myosin-I beta	+
23	**HK9**	Homo sapiens ankyrin repeat-containing protein (G9A), mRNA	+
24	**Pr16**	Insulin-like growth factor binding protein 2	-
25	**H3**	Macrophage inhibitory cytokine-1 (MIC-1)	+
26	**A22**	Cyclin-dependent kinase inhibitor 1A (p21, Cip1)	+
27	**R9**	Cyclin-dependent kinase inhibitor 1A (p21, Cip1)	+
28	**N39**	Homo sapiens lymphotoxin alpha	-
29	**N59**	Homo sapiens tumor necrosis factor	-
30	**N76**	Excision repair cross complementation group 1	-

+++ high, ++ medium, +low and - no expression of the genes

effective change was seen with high-expressing genes, for which the highest signal intensity was about twice the baseline value (Fig. 4b). The effect on signal intensity ratio for increasing concentration of the oligonucleotide probes in Table 2 reflects the minimal changes observed both for low- as well as high-expressing genes. Thus, the concentration of the probe does not seem to have a significant effect on the signal intensity, although even a three-fold increase in signal for low-expressing gene may be beneficial for detection.

Although shorter probes have the advantage of higher specificity, increasing signal with increasing length of the probe did not jeopardize the specificity. This is clear when we look at the signal from genes that show no expression in the target cell line. For example genes with ID 4 or 24 (which can be considered as negative controls) show no increased signal for increasing length or increasing concentration thus the observed increasing signal with

length or concentration of probes is specific and is not due to cross hybridization.

In this study we observed that the response to length of the oligonucleotide depends upon the level of expression of the gene of interest in the target sample. If one looks at the attachment chemistry on poly-L-lysine coated slides, the positive charge of amines at neutral pH allows attachment of native DNA or oligonucleotides through the formation of ionic bonds with the negatively charged phosphate backbone. This electrostatic attachment is supplemented by treatment with ultraviolet light or heat and induces covalent attachment of the DNA to the surface. The combination of electrostatic binding and covalent attachment couples the DNA to the substrate in a highly stable manner. Based on this, binding affinity to the surface for different length DNA is not dependent on the length. Hence the observed effect of length on signal intensity is not due to varying affinities. We conducted control experiments wherein 5' Cy3 labeled
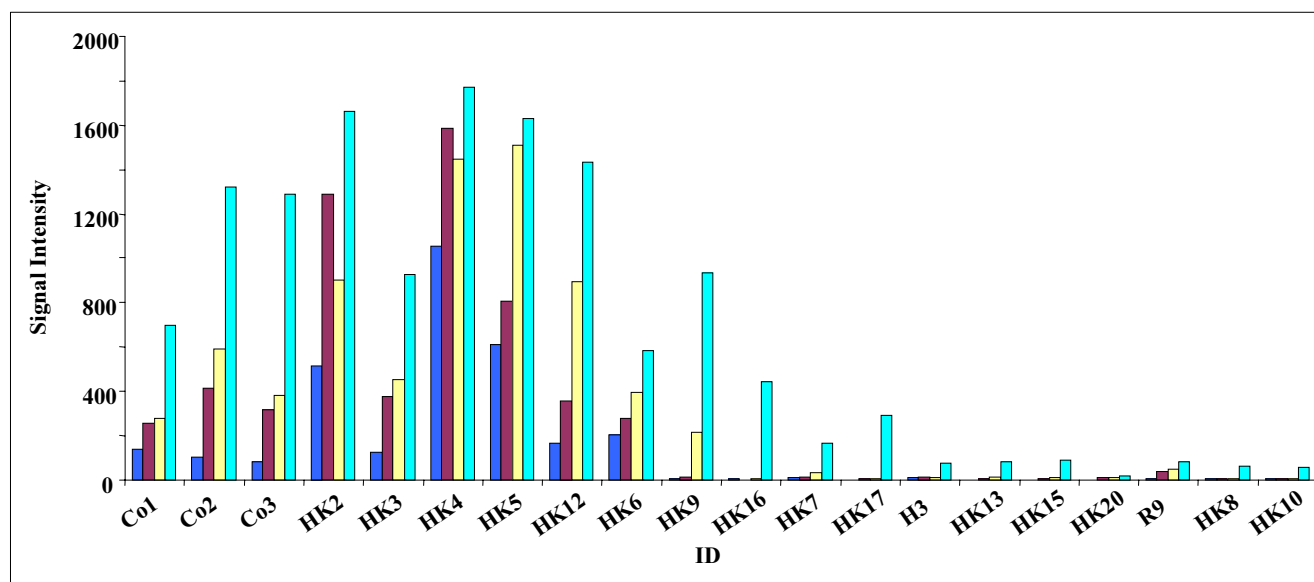
Figure 2
Absolute signal intensities for the four probe lengths representing the expression of the probed genes in the target sample. For each gene evaluated, the signal intensities from the four lengths of oligonucleotide probes are plotted on the bar graph (■ - 30 base, ■ - 40 bases, ☐ - 50 bases, ■ - 70 bases).

oligonucleotides of different length were deposited on the slide for genes of different expression levels. The relative signal intensites were measured pre and post hybridization and found not to change with increasing length of the oligonucleotides (data not shown) supporting the fact that poor attachment is not one of the reasons for the observed effect. This has also been elegantly shown by Stillman and Tonkinson [10] for varying lengths of DNA in the range 100–2000 bp, wherein they found that the K_d, the equilibrium dissociation constant for hybridization for the solution phase probe to each of the immobilized species, were all in the same range. The kinetics of hybridization will depend on the availability of the nucleotides on the probe and thus on the length of the oligonucleotide.

The signal response to increasing binding of fluorescent labeled target molecules can be represented as a binding curve and this has been shown with dilution experiments in Ramdas et. al [17]. The highly expressed genes fall in upper, more flat region of the binding curve, while the low expressers are in the most linear response region. Thus as the hybridization increases with increased length, the response is more prominent for the low expressers than for high expressers which are already in the more flat region.

Conclusions

In summary, our evaluation demonstrated that longer oligonucleotides are especially beneficial for detecting low-expressing genes. Considering that these genes are the ones most difficult to detect accurately, long oligonucleotide microarrays with 70- base nucleotides are the best option. Longer probes might not provide additional benefit because the current limitations in oligonucleotide synthesis efficiency lead to loss in full-length oligonucleotide synthesis [13,14]. Compared to the length effect, increasing the probe concentration has less dramatic effect on signal intensities for both low- and high-expression genes. Although many other features of microarray experiments influence their performance, the effect of oligonucleotide probe length on the signal intensities of low-expressing genes can clearly be controlled and optimized. Longer oligonucloeotides improve the signal for low expressing genes.

Methods
Oligonucleotide synthesis
Thirty genes with different levels of expression in the RKO colon cancer cell line were chosen for this study. The expected expression levels were based on multiple microarray data from RKO cell line. Among the 30 genes, eight were high-expressing genes, ten were low-expressing genes, three were medium-expressing genes and rest of
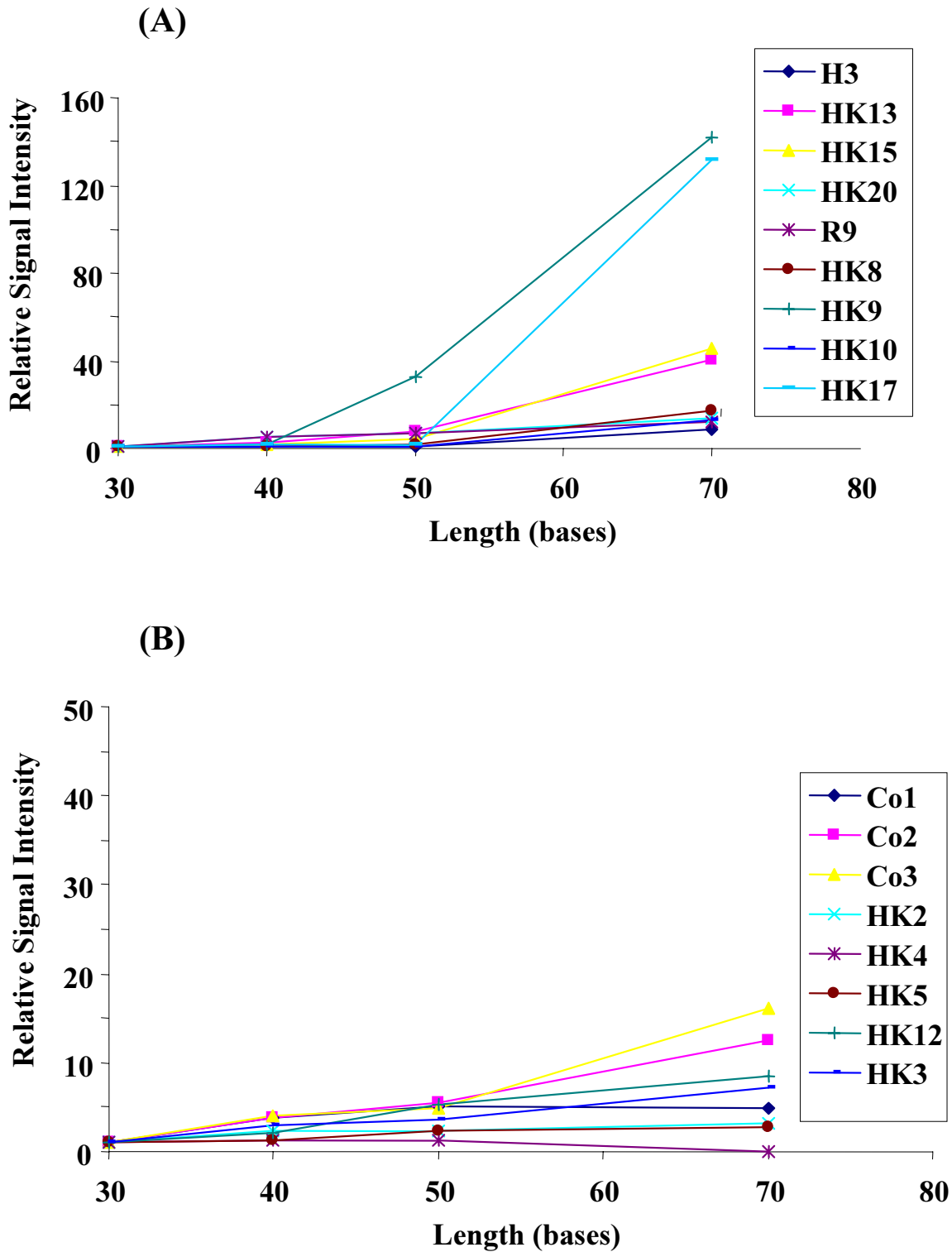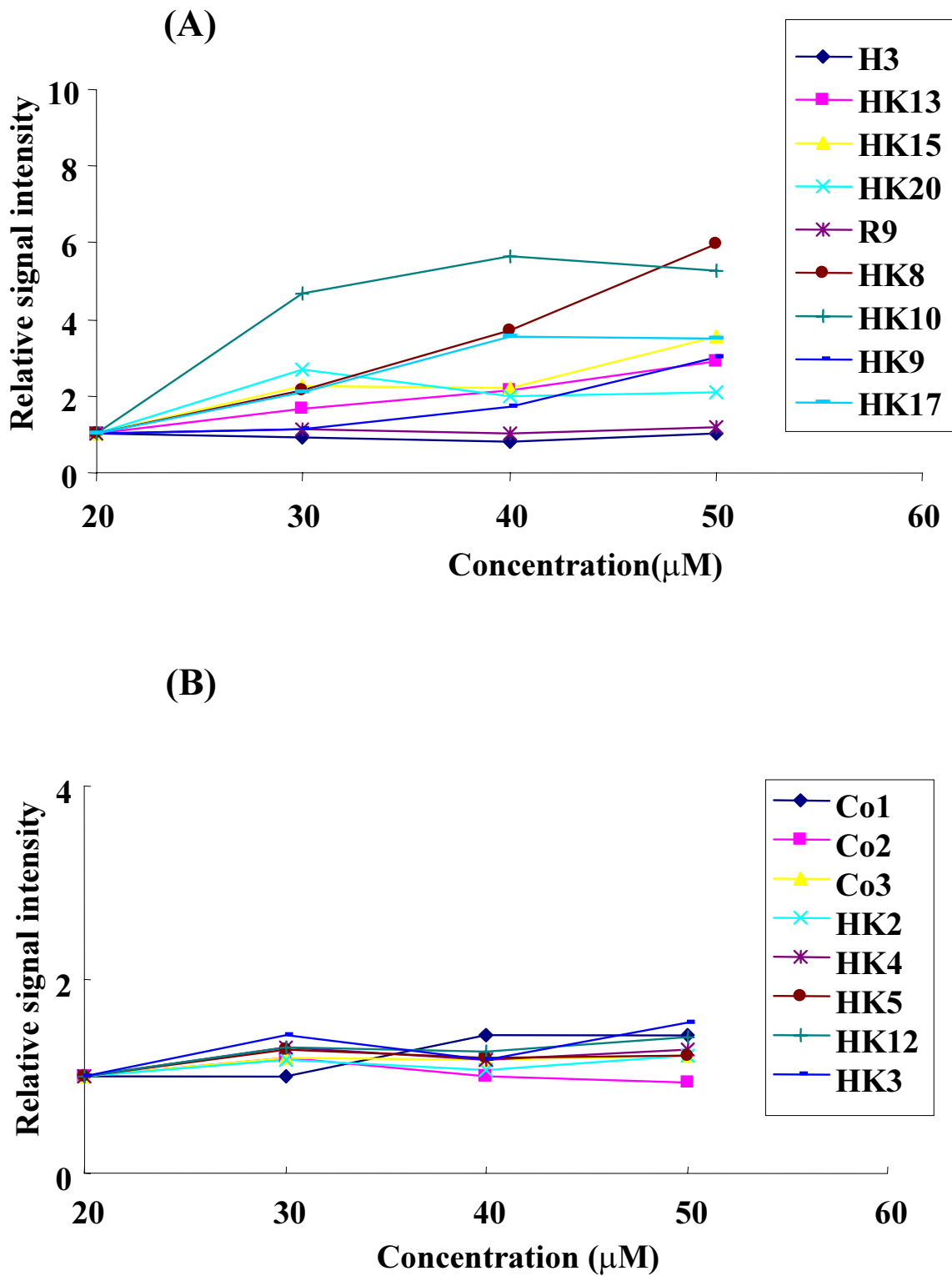
(A)

(B)

Figure 3
(A) For selected low-expressing genes (HK3, HK8, HK9, HK10, HK13, HK15, HK17, HK20, and R9), plots of the relative increase in signal intensity of the 30-, 40-, 50- and 70-base nucleotide probes in comparison with the 30-base nucleotide probe signal. The y-axis is the relative increase in signal intensity and the x-axis is the length of the probes. (B) Corresponding plots for the eight high-expressing genes (Co1, Co2, Co3, HK2, HK3, HK4, HK5 and HK12).

Figure 4
(A) For selected low-expressing genes (HK3, HK8, HK9, HK10, HK13, HK15, HK17, HK 20 and R9), plots of the relative increase in signal intensity of the 20, 30, 40 and 50 µM probe concentration in comparison with the 20 µM nucleotide probe. The y-axis is the relative increase in signal intensity and the x-axis is the concentration of the probes. (B) Corresponding plots for the eight high-expressing genes (Co1, Co2, Co3, HK2, HK3, HK4, HK5 and HK12).

Table 2: The effect of oligonucleotide length and concentration on average signal intensity ratios.

Experimental Condition		Average Intensity Ratio	
		Low-expressing genes*	High-expressing genes*
Oligonucleotide Length @ 50 μM	70	22.1 ± 13.6	7.4 ± 5.4
	50	6.1 ± 3.9	2.7 ± 1.2
	40	2.4 ± 1.3	2.0 ± 0.8
Oligonucleotide Concentration (μM) for length 70	50	3.1 ± 1.8	1.2 ± 0.2
	40	1.9 ± 1.1	0.9 ± 0.2
	30	1.5 ± 0.6	1.1 ± 0.1

* Low- and High-expressing genes are categorized based on relative expression in RKO cell line as shown in Table 1 and definitions are in results and discussion.

them showed no expression. Most of the chosen genes are considered as house keeping genes and they had varied expression in the RKO cell line. DNA sequences for the target genes were identified based on their GC content, which ranged from 35–60%, the localization of the transcript within 300 to 800 bases of the 3' end, and the minimal homology with other genes to reduce the cross-hybridization potential. The gene ID, description, and relative expression levels in the in RKO cell line are shown in Table 1. For each gene, four lengths of oligonucleotides (30, 40, 50 and 70 nucleotides) were synthesized (Table 1a – see Additional File 1). The oligonucleotides were purified by means of reverse-phase cartridge purification.

Oligonucleotide microfluidic analysis
Microfluidic analysis was performed to check each oligonucleotide's quality and quantity. Samples were resuspended in water to bring their concentration up to approximately 1,000 ng/μl. From this stock, dilutions to 100 ng/μl were made and samples assayed on an Agilent 2100 Bioanalyzer (Foster City, CA). The analysis showed that the majority of the oligonucleotides were of the correct size and purity (data not shown) and the proportion was not dependent on the length of the oligonuclotide.

Oligonucleotide Printing and hybridization
The cartridge-purified, unmodified oligonucleotides were spotted onto poly L-lysine-coated glass slides using the Genomic Solutions Flexys arrayer with 48 pins (Ann Arbor, MI). The printing was carried out at four oligonucleotide concentrations (20, 30, 40 and 50 μM) in array buffer containing 50% dimethylsulfoxide. Incubating the spotted slides at 80°C for 30 min attained attachment, and cross-linking was performed using 650 μJoules ultraviolet light. Total RNA from the RKO target cell line was extracted using the RNAeasy kit (Qiagen, Valencia, CA) [16], reverse transcribed and Cyanine-5 labeled by oligo dT priming. The oligonucleotides on the slides were hybridized with Cyanine-5-labeled total RNA in ExpressHyb solution (Clontech Laboratories, Inc., Palo Alto, CA) for 16 h at 60°C in a humid chamber. After hybridization was complete, the slides were washed sequentially at 37°C in 1x SSC (150 mM sodium chloride and 15 mM sodium citrate) plus 0.01% SDS, 0.2x SSC plus 0.01% SDS, and twice in 0.1x SSC for 2 min at each step [16].

Imaging and data analysis
After the hybridization and washing steps were complete, the slides were scanned using a GeneTac LSIV laser scanner (Genomic Solutions, Ann Arbor, MI). The signal intensities were quantified using ArrayVision spot finding program (Imaging Research Inc., St. Catherines, Ontario, Canada). The signal intensities of duplicate spots for each gene at each oligonucleotide length and concentration were averaged for further analysis. Relative intensities were calculated by comparing relevant signal intensity with that at the lowest concentration (20 μM) or shortest oligonucleotide length (30 nucleotides), the values of which were set at 1.

Authors' contributions
LR designed the experiments, analyzed, interpreted the data and prepared the manuscript. DEC printed the arrays, JYJ scanned and quantified the arrays, ET, VRD and LH checked the oligonucleotides for purity, purified RNA and performed the hybridization. SRH and WZ supervised the experiments and participated in the manuscript preparation. All authors read and approved the final manuscript.

Additional material

References

1. Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* 1995, **270:**467-470.
2. Barczak A, Rodriguez MW, Hanspers K, Koth LL, Tai YC, Bolstad BM, Speed TP, Erle DJ: **Spotted long oligonucleotide arrays for human gene expression analysis.** *Genome Res* 2003, **13:**1775-1785.
3. Taylor E, Cogdell D, Coombes K, Hu L, Ramdas L, Tabor A, Hamilton SR, Zhang W: **Sequence verification as quality-control step for production of cDNA microarrays.** *Biotechniques* 2001, **31:**62-65.
4. Evertsz EM, Au-Young J, Ruvolu MV, Lim AC, Reynolds MA: **Hybridization cross-reactivity within homologous gene families on glass cDNA microarrays.** *Biotechniques* 2001, **31:**1182-1192.
5. Relogio A, Schwager C, Tichter A, Ansorge W, Valcarcel J: **Optimization of oligonucleotide-based DNA microarrays.** *Nucleic Acid Res* 2002, **30:**e51.
6. Kumar A, Larssaon O, Parodi D, Liang Z: **Silanized nucleic acids: a general platform for DNA mobilization.** *Nucleic Acids Res* 2000, **28:**e71.
7. Wang H, Malek RL, Kwitek AE, Greene AS, Luu TV, Behbahani B, Frank B, Quackenbush J, Lee NH: **Assessing unmodified 70-mer oligonucleotide probe performance on glass-slide microarrays.** *Genome Biology* 2003, **4:**R5.
8. **Nucleotide BLAST** [http://www.ncbi.nlm.nih.gov/BLAST/]
9. Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, Shannon KW, Lefkowitz SM, Ziman M, Schelter JM, Meyer MR, Kobayashi S, Davis C, Dai H, He YD, Stephaniants SB, Cavet G, Walker WL, West A, Coffey E, Shoemaker DD, Stoughton R, Blanchard AP, Friend SH, Linsley PS: **Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer.** *Nature Biotechnology* 2001, **19:**342-347.
10. Stillman BA, Tonkinson JL: **Expression microarray hybridization kinetics depend on length of the immobilized DNA but are independent of immobilization substrate.** *Anal Biochem* 2001, **295:**149-157.
11. Verma S, Eckstein F: **Modified Oligonucleotides: Synthesis and strategy for users.** *Ann Rev Biochem* 1998, **67:**99-134.
12. Mir KU: **What length probe is optimal for microarrays?** *Nature Genetics* 1999, **23:**63. Poster Abstracts
13. [http://www.annovis.com/technotes/technote01.pdf].
14. [http://dna-9.int-med.uiowa.edu/OligoFAQs.htm].
15. Kane MD, Jatkoe TA, Stumpf CR, Lu J, Thomas JD, Madore SJ: **Assessment of the sensitivity and specificity of oligonucleotides (50 mer) microarrays.** *Nucleic Acids Res* 2000, **28:**4552-4557.
16. Hu L, Wang J, Baggerly K, Wang H, Fuller GN, Hamilton SR, Coombes KR, Zhang W: **Obtaining reliable information from minute amounts of RNA using cDNA microarrays.** *BMC Genomics* 2002, **3:**16-24.
17. Ramdas L, Coombes KR, Baggerly K, Hess K, Abrruszzo L, Zhang W: **Sources of nonlinearity in cDNA microarray expression measurements.** *Genome Biol* 2001, **2:**1-47.