

## Application of hybridization control probe to increase accuracy on ligation detection or minisequencing diagnostic microarrays

Jarmo Ritari\*, Lars Paulin, Jenni Hultman and Petri Auvinen

Address: DNA sequencing and genomics laboratory, Institute of Biotechnology, University of Helsinki, 00790 Helsinki, Finland

Email: Jarmo Ritari\* - jarmo.ritari@helsinki.fi; Lars Paulin - lars.paulin@helsinki.fi; Jenni Hultman - jenni.hultman@helsinki.fi; Petri Auvinen - petri.auvinen@helsinki.fi

\* Corresponding author

Published: 14 December 2009

Received: 22 June 2009

BMC Research Notes 2009, 2:249 doi:10.1186/1756-0500-2-249

Accepted: 14 December 2009

This article is available from: <http://www.biomedcentral.com/1756-0500/2/249>

© 2009 Ritari et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Nucleic acid detection based on ligation reaction or single nucleotide extension of ssDNA probes followed by tag microarray hybridization provides an accurate and sensitive detection tool for various diagnostic purposes. Since microarray quality is crucial for reliable detection, these methods can benefit from correcting for microarray artefacts using specifically adapted techniques.

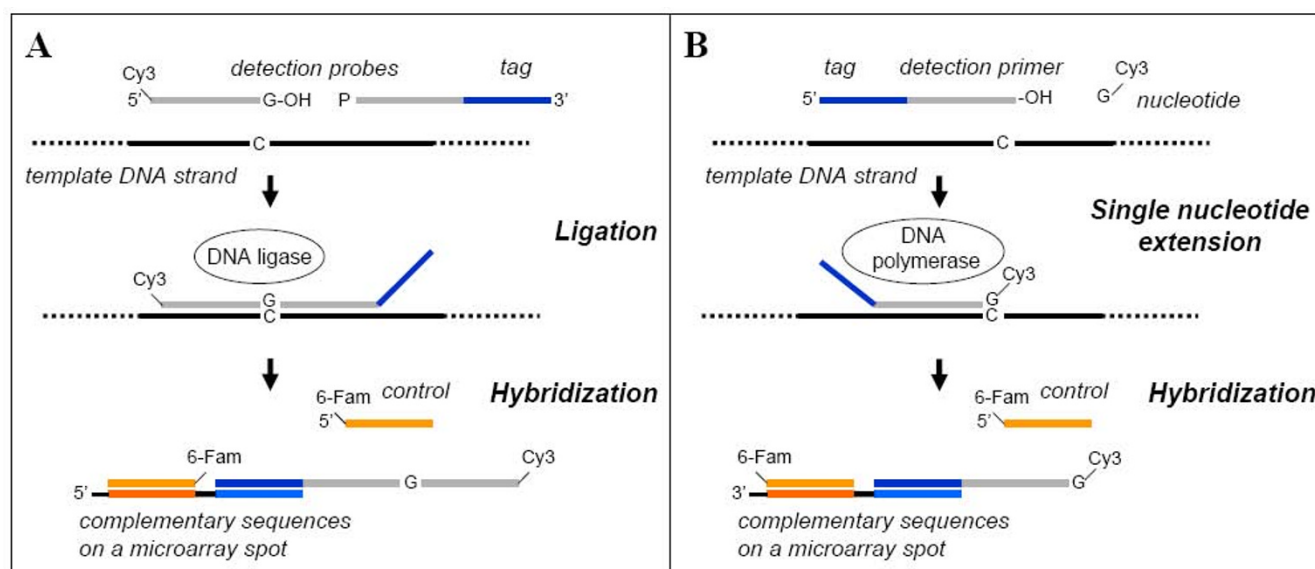
**Findings:** Here we demonstrate the application of a per-spot hybridization control oligonucleotide probe and a novel way of computing normalization for tag array data. The method takes into account the absolute value of the detection probe signal and the variability in the control probe signal to significantly alleviate problems caused by artefacts and noise on low quality microarrays.

**Conclusions:** Diagnostic microarray platforms require experimental and computational tools to enable efficient correction of array artefacts. The techniques presented here improve the signal to noise ratio and help in determining true positives with better statistical significance and in allowing the use of arrays with poor quality that would otherwise be discarded.

### Background

Nucleic acid detection by ligation and single-nucleotide extension minisequencing techniques take advantage of the catalytic selectivity of DNA ligase and polymerase enzymes, respectively, to recognize a unique position in a target DNA strand. In ligation assays, two specific ssDNA oligonucleotide detection probes are designed to hybridize adjacently on target DNA strand so that the 3' end of the label-carrying probe recognizes a discriminating position and is ligated to the phosphorylated 5' end of the other probe in the presence of a matching target molecule (figure 1A) [1,2]. Ligation detection can also be imple-

mented as a single probe which is circularized upon ligation [3]. In minisequencing, the target is recognized through the addition of a specific labeled dideoxynucleotide to the 3' end of the oligonucleotide detection primer annealed immediately upstream of a discriminating position in the target (figure 1B) [4,5]. Both methods allow tagging of the probes for detection on a microarray platform containing complementary tag sequences providing uniform thermodynamic hybridization properties for all probes. The relatively high throughput and superior accuracy over traditional microarray and PCR based methods have motivated the application of ligation and minise-

**Figure 1**

**Principle of ligation detection and minisequencing reactions.** A) Two ssDNA probes with one carrying a 5' fluorescent label and the other having a 5' phosphorylation and a 3' flanking tag sequence hybridize adjacently on target DNA. DNA ligase accepts nicked dsDNA strand as a substrate and joins the probes together if there is a perfect match to the template at the 3' end of the probe. B) Minisequencing is a similar approach using a detection primer and single dideoxynucleotide extension catalysed by DNA polymerase. The tag is in the 5' end of the primer and the nucleotide carries a fluorescent label incorporated into the 3' end. The reaction products are hybridized onto a microarray with complementary tag sequences. Each spot also harbors a complementary sequence to a 5' fluorescently labeled control probe which is read on another wavelength channel.

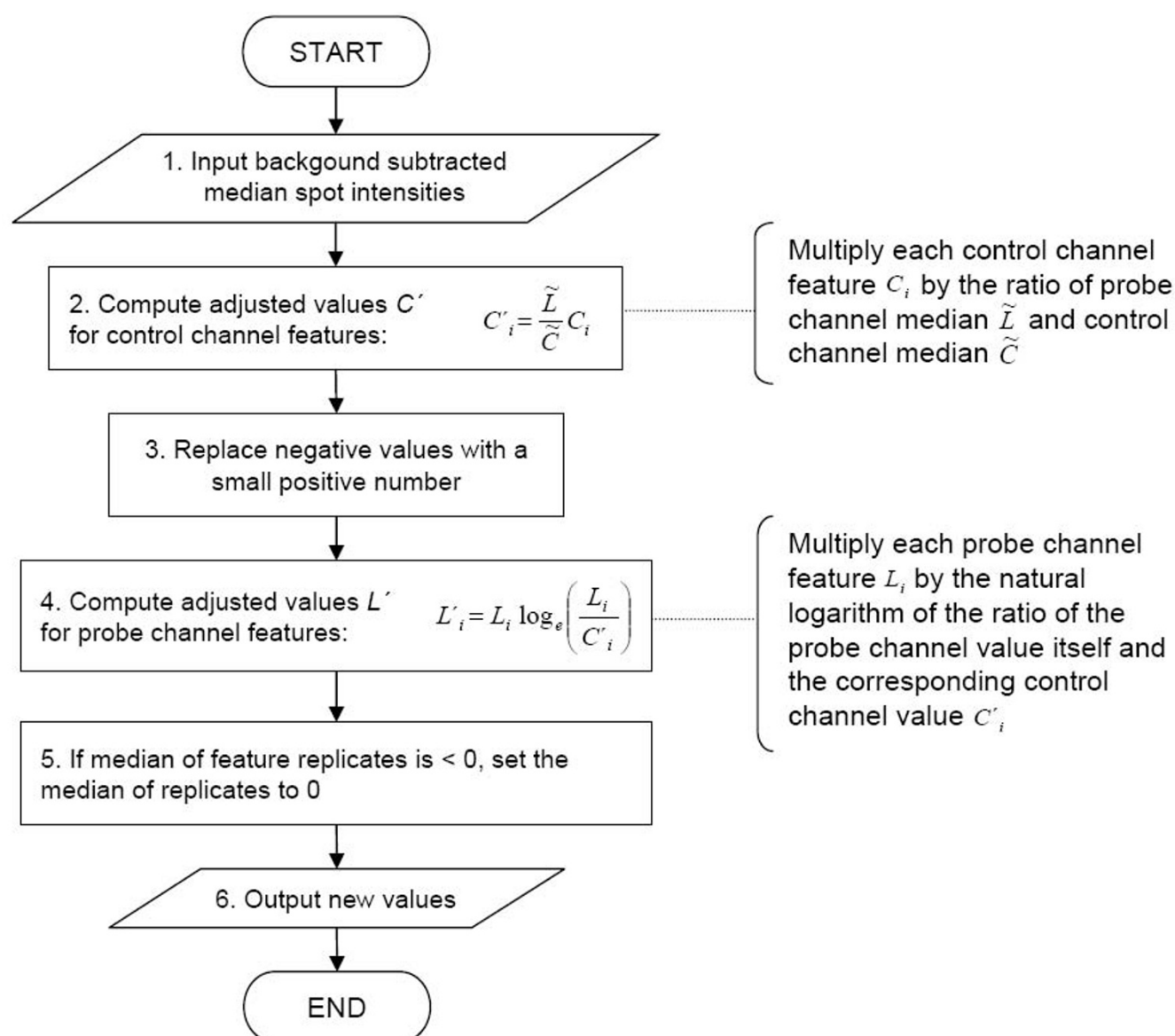
quencing probe microarrays to SNP [6,7] and gene variant detection [8], clinical microbial diagnostics [9-11] and more recently to environmental microbiology [12-14]

Even though enzyme aided recognition provides good accuracy as such, ultimately sensitivity and reliability are dependent on microarray quality and successful hybridization. The fidelity of recorded intensities of hybridized detection probes might be adequate for diagnostics provided the array spot quality is constantly good throughout the array as is often the case with *in situ* synthesized and other high quality microarrays. However, aberrant spots that vary in morphology or DNA content can occur for instance in contact printed arrays, causing problems with accuracy of spot finding and quantification. In addition, the printing process can introduce additional background noise impeding the read-out of results. Therefore, diagnostic microarrays may benefit from information processing steps to remove biases and noise, but this requires additional experimental measures to determine the source of variance. Methods used in gene expression normalization are not directly applicable because they typically rely on the bulk signal of all array spots assuming only that a small minority of genes are differentially expressed (reviewed in [15,16]). In diagnostic microarrays, this assumption does not generally hold and also in some applications the number of spots per array is much lower.

## Method

Here we report an approach using a hybridization control oligonucleotide probe to measure tag array quality independently of detection probes (figure 1) and to enable normalization in order to remove noise and standardize signals between spots and subarrays (figure 2). The control probe positive for all spots is similar to the tag oligos with regard to length and base composition. The detection probe signal intensities are compared to the control probe signal intensities for each spot to obtain normalized signals. However, simply dividing the probe signals by control can be problematic if a spot has reduced intensity as a result of abnormal morphology or otherwise compromised quality. These kind of features can give rise to false positives when the probe channel is empty (due to no target present in the sample) because the background signal in the probe channel is relatively constant and independent of the quality of the spot. The control signal, on the other hand, reflects the spot quality much more closely because the control probe is positive to all spots. Thus, division of low but constant detection probe background signal by reduced control signal may produce artificially high ratio values (an example is given in figure 3).

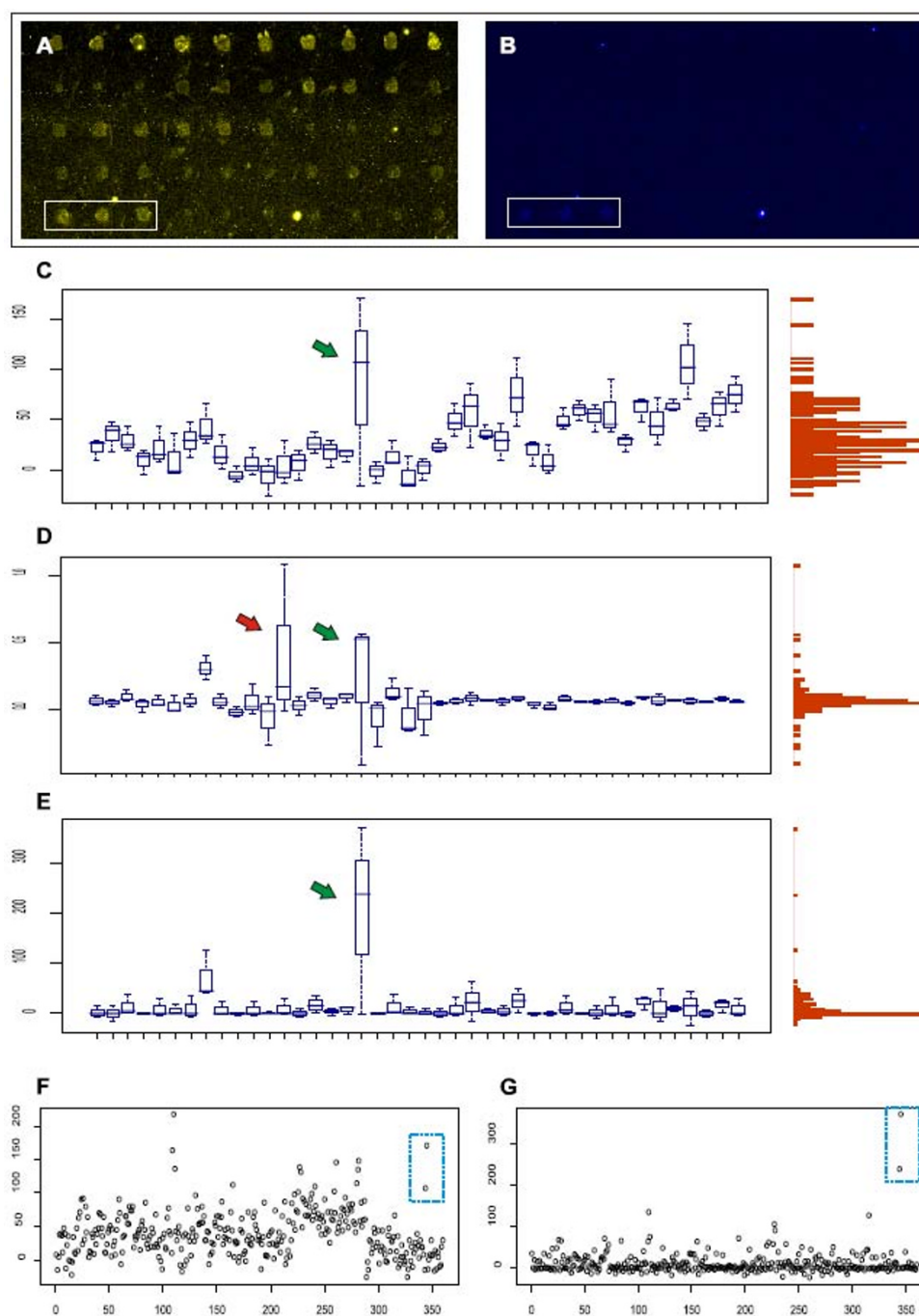
To avoid generating false positives in the analysis, we have first computationally adjusted the control channel intensities to the level of detection probe channel median of all

**Figure 2**

**Flow diagram describing computational steps of the normalization algorithm.** Input data are median background subtracted median spot pixel intensities. In the second step, control channel spot intensities are adjusted to the level of probe channel intensities. Next, negative values are replaced by a small number to avoid taking negative logarithm. Fourth, probe channel spot values are adjusted by multiplying the probe spot value by natural logarithm of the corresponding ratio of probe and control spot values. Finally, if median of replicate spots is negative, the difference is added to the spot values to set the median to the minimum of zero.

spots on a subarray in order to prevent the typically stronger control probe signal from dominating over the detection probe signal. The assumption behind this is that the median value represents empty detection probe signal. Next, the probe to control ratios are computed and used in logarithm as weighting coefficients in computing adjusted values for detection probe spot signals (figure 2). The advantage of this approach is that the probe signal

value is multiplied by the weighting coefficient value over zero only if the detection probe signal is stronger than the control in that particular spot. If the spot is empty in the detection probe channel, the result of multiplication by the log ratio gives a small number even if the log ratio would be positive, unlikely to cause any false calls. In addition, as the log ratio and consequently the output is 0 with equal probe and control values, it provides a com-

**Figure 3**

**An example of a microarray with high background noise and poor spot quality.** Plots showing results before and after applying the normalization procedure. A) and B) show a part of a microarray scanned in two channels; wavelength 488 (control with 6-Fam dye) and wavelength 532 (probes with Cy3 dye), respectively. The marked spot triplet in the bottom left corner shows a positive probe. Boxplots in C) show hybridization results of 42 detection probes with a positive probe indicated by green arrow. The probes are listed on the x-axis and their relative intensities on the right y-axis. On the left y-axis, a vertical histogram depicting the intensity distribution. The values are background subtracted detection probe channel signal intensities. D) shows the ratios of detection probe to adjusted control. Red arrow points a false positive. In E), the normalization described here is applied to the data, making positive signals better defined. F) and G) show all spots in the array before and after normalization, respectively. Marked spots in the top right corner indicate the positive probe.

mon reference point for all spots within and between microarrays.

### Implementation

The probe set used in the normalization experiments consisted of ligation detection probes similar to a previous study [13]. Forty-two different probes (the sequences are to be published elsewhere) were multiplexed in ligation reactions. The ligation reactions and hybridization conditions were as described previously [13]. Briefly, the ligations were carried out in a final volume of 20  $\mu$ l containing 1 $\times$  ligation buffer (TAQ ligase buffer, New England Biolabs, MA, USA), 30 mM tetramethylammonium chloride (TMAC), 250 fmol of each discriminating probe, 250 fmol of each common probe, 5 pmol of the complementary hybridisation control probe, a variable amount of purified PCR products and 4 U of Taq DNA ligase (New England Biolabs). The reaction was cycled for 40 rounds at 94°C for 30 s and at 60-64°C for 4 min in a thermocycler (MJ Research). The LDR mix (20  $\mu$ l) was diluted to obtain 40  $\mu$ l of hybridization mixture containing 5 $\times$  SSC and 0.1 mg/ml herring sperm DNA. After heating the mix to 94°C for 2 min and chilling on ice, ligation control probe was added and the mix was applied onto the slide. The microarray slides were produced by contact printing by Telechem (CA, USA) or by university core facility (Biomedicum Biochip Center, University of Helsinki, Finland). The microarrays had 16 subarrays each, consisting of 119 tag oligos in triplicates [see Additional file 1]. Scanning were done as described previously [13]. All computations were done in R-software environment (2.8.0) [17], using the Bioconductor package Marray (1.20.0) [18] [see Additional file 2] for reading GenePix result (gpr) files [see Additional file 3]. In order to evaluate the effect of normalization, no filtering or outlier removal procedures were applied to the data.

### Results

The normalization procedure was tested with two different sets of detection probes on microarrays containing either three or five spot replicates. Some subarrays had printing artefacts like background noise or low quality spots. Figure 3 shows an example of a typical poor quality subarray. In these kind of arrays, the spot morphology varied considerably and background noise was high in places resulting in signal-to-noise ratio much lower than in regular microarrays. Comparison of results before and after applying the control probe and normalization demonstrates the impact of these procedures on the signal-to-noise ratio. Clearly, just subtracting background signal from each spot raw value in the detection probe channel is not enough to provide reliable results (figure 3C). Computing signal ratio of the detection probe to the adjusted control (figure 3D) greatly reduces variation in the background distribution but at the same time weakens true

positive signal and causes some spots to falsely reach too high values. The normalization procedure presented here, however, avoids these pitfalls and is able to correct for the variation and keep the true positive signal clear (figure 3E and 3G). We tested 8 microarray slides with altogether 128 subarrays. Furthermore, the procedure was applied on data from previously published microarray hybridizations [see Additional file 4]. In this set, the ligation probes were designed against environmental fungi [13] and detected on a tag microarray with five spot replicates. Also in this case the normalization was capable of correcting microarray noise. As expected, the computations had little or no effect on results from good quality microarrays [see Additional file 5].

### Discussion

DNA microarrays, while being a potential tool for diagnostics, typically have irregularities in spot quality causing problems with accuracy of detection. Efficient background correction is required in ligation, minisequencing and similar diagnostic systems where standard gene expression analysis methods may not apply. Although high quality microarrays are preferable for diagnostics, they may not always be economically feasible for routine use due to high costs. The prices of customizable commercial high-density *in situ* synthesized microarrays are likely decrease but these kind of platforms can still benefit from noise correction as human error in performing the hybridizations can not be ruled out. In addition, some emerging diagnostic platforms such as integrated microfluidic systems can take advantage of microarrays for detection of probes on a small scale [19,20]. The mass fabrication process of the devices and hybridization conditions might bring about variation in spot quality and available space for replicates in microsystems is likely to be limited as well. In diagnostic applications in general, accuracy of the results is highly important and proper correction procedures can help deal with noise to increase statistical reliability, making the system practicable even if the detection platform is not fully optimal.

We found the normalization procedure based on an internal control for each assay spot to be useful when working with tag microarrays having compromised spot quality and background noise. A similar approach has been used by others before to monitor array spot quality and to correct for printing variations. In a study by Ye and coworkers, amplicons were first generated by forward and reverse primers carrying different labels [21]. One strand of a amplicon served as target for SNP recognition probes while the other strand was used as a internal control hybridizing to a complementary control probe in the same spot. In another work, a 25 mer control oligo was spotted alongside with 70 mer recognition probes to monitor spot quality with a complementary labeled 25

mer [22]. In both of these studies, a given spot control signal was first compared with the mean control signal and the obtained value was then used to divide the detection signal of the same spot. Similarly, Akhras and coworkers computed signal ratios of detection probe vs. an all-positive control on a padlock probe tag array [9]. However, problematic microarrays or abnormal spots were not the focus of these studies and correcting microarray defects was not discussed in further detail. The method presented here uses a similar principle of internal control but different computational procedure to effectively reduce noise on low quality spots, emphasizing signal extraction rather than mere elimination of problematic spots. It is important not only to monitor the array spot quality but also to take into account the possible bias in probe to control ratio to effectively process aberrant spots on low quality arrays.

We have also demonstrated earlier with cDNA microarrays that effective measurement of spot quality with an additional dye improves the reliability of detection [23]. Using an all-positive control probe serves a similar idea in assisting spot quantification which is highly dependent on the accurate determination of the true spot area to estimate spot and background signal intensities. This is especially relevant if the signal intensity of a spot is low in the detection channel. The control channel, having a relatively high intensity in all spots, helps in locating the spots and capturing their areas more accurately. However, it should be noted that our method does not use all the information available in the control channel. For instance, the control signal could be used to model the detection probe channel intensity profile. This approach could potentially increase the signal-to-noise-ratio of the detection probe channel on weak spots, opening the possibility for further development of the method.

The computation presented here can be easily implemented in any freely programmable system used for microarray data analysis. It should be noted however, that application of the control probe requires that each spot on the tag microarray harbors the complementary control sequence, along with the actual tag sequence. One way to overcome the need to synthesize novel long probes is to mix control oligonucleotide serving as an internal control with each of the specific oligonucleotides and deposit these mixture on arrays [22]. In addition, if it is expected that over 50% of detection probes should be positive in an experiment, negative control spots are needed to compute the normalization. This is because the procedure assumes that the detection probe channel median represents empty signal in computing adjusted signal values for the control channel.

## Conclusions

Much of the noise introduced by variable spot quality on detection probe read-outs can be corrected by applying simple computations. This involves adjusting the control probe signals to detection probe signal levels and taking into account absolute value of the detection probe signal and variability in the control probe signal. The method is potentially advantageous in various diagnostic microarray platforms.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

JR analysed the data and wrote the scripts. JR, LP, JH and PA conceived of the study and drafted the manuscript. All authors read and approved the final manuscript.

## Additional material

### Additional file 1

**Tag\_sequences.** Sequences of tags printed on microarrays and complementary tags flanking the detection probes.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1756-0500-2-249-S1.XLS>]

### Additional file 2

**Scripts.** R file containing scripts for reading data, computing normalization and drawing boxplots.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1756-0500-2-249-S2.RTF>]

### Additional file 3

**Gpr\_files.** Zip-compressed file containing raw microarray data in GenePix results format (.gpr) and GAL (GenePix Array List) files.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1756-0500-2-249-S3.ZIP>]

### Additional file 4

**Supplemental\_figure1.** Boxplots showing data from a different probeset published previously [13]. The probes are listed on the x-axis and their relative intensities on the right y-axis. On the left y-axis, a vertical histogram depicting the intensity distribution. Signals before (A) and after (B) the normalization procedure. The probes are designed to detect different environmental fungi and were hybridized on microarray with tag sequences spotted in 5 replicates. The normalization is capable of correcting noise and making the signal from true positive *Penicillium commune* probe (tag A50) clearer as indicated by blue arrows. For the sake of clarity, another positive probe on tag A29 was left out from the plot. Tags A100 and A101 represent empty spots (i.e. without corresponding detection probes).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1756-0500-2-249-S4.PDF>]



### Additional file 5

**Supplemental\_figure2.** Boxplots showing data from a good quality microarray before (A) and after (B) normalization. The probes are listed on the x-axis and their relative intensities on the right y-axis. On the left y-axis, a vertical histogram depicting the intensity distribution.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1756-0500-2-249-S5.PDF>]

### Acknowledgements

The Finnish Cultural Foundation is acknowledged for supporting JR financially. PA and JH were supported financially by the European regional development fund (ERDF) and Maj and Tor Nessling Foundation. We are also grateful to docent Eeva Auvinen from the Haartman Institute of the University of Helsinki for contributing ligation probes, and to Rita Fingerroos for performing microarray hybridizations.

### References

- Landegren U, Kaiser R, Sanders J, Hood L: **A ligase-mediated gene detection technique.** *Science* 1988, **241**(4869):1077-1080.
- Gerry NP, Witowski NE, Day J, Hammer RP, Barany G, Barany F: **Universal DNA microarray method for multiplex detection of low abundance point mutations.** *J Mol Biol* 1999, **292**(2):251-262.
- Nilsson M, Malmgren H, Samiotaki M, Kwiatkowski M, Chowdhary BP, Landegren U: **Padlock probes: circularizing oligonucleotides for localized DNA detection.** *Science* 1994, **265**(5181):2085-2088.
- Syvänen AC, Aalto-Setälä K, Harju L, Kontula K, Soderlund H: **A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E.** *Genomics* 1990, **8**(4):684-692.
- Pastinen T, Kurg A, Metspalu A, Peltonen L, Syvänen AC: **Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays.** *Genome Res* 1997, **7**(6):606-614.
- Sigurdsson S, Hedman M, Sistonen P, Sajantila A, Syvänen AC: **A microarray system for genotyping 150 single nucleotide polymorphisms in the coding region of human mitochondrial DNA.** *Genomics* 2006, **87**(4):534-542.
- Hardenbol P, Yu F, Belmont J, Mackenzie J, Bruckner C, Brundage T, Boudreau A, Chow S, Eberle J, Erbilgin A, Falkowski M, Fitzgerald R, Ghose S, Iartchouk O, Jain M, Karlin-Neumann G, Lu X, Miao X, Moore B, Moorhead M, Namsaraev E, Pasternak S, Prakash E, Tran K, Wang Z, Jones HB, Davis RW, Willis TD, Gibbs RA: **Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay.** *Genome Res* 2005, **15**(2):269-275.
- Zeng F, Ren ZR, Huang SZ, Kalf M, Mommersteeg M, Smit M, White S, Jin CL, Xu M, Zhou DW, Yan JB, Chen MJ, van Beuningen R, Huang SZ, den Dunnen J, Zeng YT, Wu Y: **Array-MLPA: comprehensive detection of deletions and duplications and its application to DMD patients.** *Hum Mutat* 2008, **29**(1):190-197.
- Akhras MS, Thiyagarajan S, Villablanca AC, Davis RW, Nyren P, Pourmand N: **PathogenMip assay: a multiplex pathogen detection assay.** *PLoS ONE* 2007, **2**(2):e223.
- Kostic T, Weilharter A, Rubino S, Delogu G, Uzzau S, Rudi K, Sessitsch A, Bodrossy L: **A microbial diagnostic microarray technique for the sensitive detection and identification of pathogenic bacteria in a background of nonpathogens.** *Anal Biochem* 2007, **360**(2):244-254.
- Szemes M, Bonants P, de Veerdt M, Baner J, Landegren U, Schoen CD: **Diagnostic application of padlock probes--multiplex detection of plant pathogens using universal microarrays.** *Nucleic Acids Res* 2005, **33**(8):e70.
- Busti E, Bordoni R, Castiglioni B, Monciardini P, Sosio M, Donadio S, Consolandi C, Rossi Bernardi L, Battaglia C, De Bellis G: **Bacterial discrimination by means of a universal array approach mediated by LDR (ligase detection reaction).** *BMC Microbiol* 2002, **2**:27.
- Hultman J, Ritari J, Romantschuk M, Paulin L, Auvinen P: **Universal ligation-detection-reaction microarray applied for compost microbes.** *BMC Microbiol* 2008, **8**:237.
- Rantala A, Rizzi E, Castiglioni B, de Bellis G, Sironen K: **Identification of hepatotoxin-producing cyanobacteria by DNA-chip.** *Environ Microbiol* 2008, **10**(3):653-664.
- Quackenbush J: **Computational analysis of microarray data.** *Nat Rev Genet* 2001, **2**(6):418-427.
- Quackenbush J: **Microarray data normalization and transformation.** *Nat Genet* 2002, **32**(Suppl):496-501.
- R Development Core Team: **R: A language and environment for statistical computing.** 2008 [<http://www.R-project.org>]. R Foundation for Statistical Computing, Vienna, Austria ISBN 3-900051-07-0
- Yee Hwa (Jean) Yang with contributions from Agnes Paquet and Sandrine Dudoit: **marray: Exploratory analysis for two-color spotted microarray data.** R package version 1.20.0; 2007.
- Hashimoto M, Hupert ML, Murphy MC, Soper SA, Cheng YW, Barany F: **Ligase detection reaction/hybridization assays using three-dimensional microfluidic networks for the detection of low-abundant DNA point mutations.** *Anal Chem* 2005, **77**(10):3243-3255.
- Hashimoto M, Barany F, Soper SA: **Polymerase chain reaction/ligase detection reaction/hybridization assays using flow-through microfluidic devices for the detection of low-abundant DNA point mutations.** *Biosens Bioelectron* 2006, **21**(10):1915-1923.
- Yin BC, Li H, Ye BC: **A dual-probe hybridization method for reducing variability in single nucleotide polymorphism analysis with oligonucleotide microarrays.** *Anal Biochem* 2008, **383**(2):270-278.
- Peterson G, Bai J, Narayanan S: **A co-printed oligomer to enhance reliability of spotted microarrays.** *J Microbiol Methods* 2009, **77**(3):261-266.
- Gupta R, Ruosaari S, Kulathinal S, Hollmen J, Auvinen P: **Microarray image segmentation using additional dye--an experimental study.** *Mol Cell Probes* 2007, **21**(5-6):321-328.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

