Microarray probe set: Biology, bioinformatics and biophysics

Payam Behzadi[1], Elham Behzadi[2], Reza Ranjbar[1]

[1]Molecular Biology Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran;
[2]Academy of Medical Sciences of the Islamic Republic of Iran, Tehran, Iran.

Corresponding author: Dr. Reza Ranjbar;
Address: Shahid Nosrati alley, Sheikh Bahaee Avenue, Molla Sadra Street, Vanak Square, Tehran, Iran;
Telephone: +982188039883; E-mail: ranjbarre@gmail.com

Abstract

Aim: In spite of improvements in microarray platforms, there are some uncertainties, biases and errors in data outcomes and ultimate data interpretations. One of the most considerable problems in microarray methodologies is the choice of probe sets. For selecting an appropriate probe set, designing and calibrating the related probes, there is an unavoidable need to microarray methodologies, bioinformatics, biophysics and their relationship. Therefore, cooperation and collaboration between biologists, bioinformaticians and biophysicists is a must. For this reason, the aim of this literature review is to show the multi-dimensional effects of probes in final analytical outcomes throughout array designing, probe sets, probe calibration, noise calculation, and hybridization.
Methods: A good choice of probe set in association with accurate probe calibration and minimum noise production is needed as an acceptable and prompt methodology to decrease the erroneous of data analysis and biased interpretation.
Results: The use of interdisciplinary sciences including bioinformatics, biophysics, and biology for designing an accurate probe set in microarray technology ensures the quality of data analysis and unbiased interpretation. Hence, a proper methodology for choosing an appropriate probe set improves the reliability of the results obtained via the microarray technique.
Conclusion: The application of suitable probe sets results in precise and accurate protocols. The occurrence of this progression is related to an intense collaboration between biophysicists, bioinformaticians and biologists.

Keywords: bioinformatics, biophysics, microarray, probe set.

Introduction

There are several significant and invaluable biological and medical applications including Gene Expression Profiling, Genetic Disorders Diagnosis, and Infectious Diseases Diagnosis, and Pathogenic Microorganisms Detection which may be done via microarrays (1).

The progressions in different basic science disciplines involving biology, biochemistry, computer, genetics, mathematics, molecular biology and interdisciplinary subjects such as bioinformatics and biophysics provide the accessibility to a huge amount of data by modern technologies and approaches. Among a diversity of modern and specific tools and techniques, microarrays are known as fluorescent nucleic acid based technologies which are associated with labeled probes linked on a chip. Microarrays are miniaturized and accelerated automatic techniques which have a variety of applications including Gene expression investigations, Genomic evolutionary rate profiling, Universal arrays, Sequence analysis, and clinical/ environmental microbial diagnostics (detection/ identification use) in bio-defense, food, medicine, and veterinary microbiology (2-5).

Microarray needs a huge amount of nucleic acids as a target sample, but on the other hand the enzymatic amplifications are omitted from this technology. Therefore, these properties may reduce the technical biases in microarray (3).

Microarray technology is a collection of probes spotted on coated microscope glass slides (or carbon, metals, and silicon) via spotter, labeled target genome, hybridization, scanner and associated analytical software (1,3,5-7).

The use of prompt probe sets improves the quality of mapping and raw data which may lead to qualified data analysis and interpretation (8).

The substrates coating microarray slides involve Aldehyde Silane, Amino Silane, Epoxy Silane, hydrogel, Nitrocellulose, Plastic Poly-L-Lysine, and Thin film 3-D polymer. The aforementioned materials are general covalent binding immobilizers that capture probes including carbohydrates, nucleic acids, and proteins for attaching to target sequences in microarray diagnostic format (3,9).

Yet, despite the increase of microarray applications also known as a speed-to-answer technique, it is an expensive analytical technology with some opacities, ambiguousness and biases in the interpretation pertaining to ultimate outcomes. For this reason, there are some attempts to reduce the aforementioned complications associated with microarray technology and increase the accuracy, simplicity and unambiguousness of analytical performance (9-11).

The aim of the present literature review is to show the multi-dimensional effects of probes in final analytical outcomes throughout array designing, probe sets, probe calibration, noise calculation, and hybridization.

Array designs: Advantages and revisions

In old generations of microarray platforms, because of the presence of the noise problem relating to rough and imbalanced spotted arrays, there was a high uncertainty in the analytical performance and interpretations. Today, the noise problem and manufacturing methods are improved for the most. According to previous studies, the erroneous probe attachment performance may lead to produce measurement and interpretation errors. Thus, array designs, appropriate probe sets, accurate probe calibration and successful hybridization may help to solve unreliability associated with biased analytical interpretations and noise signals (12,13).

Probe designing in microarray technology is pure bioinformatic knowledge - an *in silico* procedure based on several software tools such as Probe (NCBI) (http://www.ncbi.nlm.nih.gov/probe/), Picky (http://www.complex.iastate.edu/download/Picky/), OligoWiz (http://www.cbs.dtu.dk/services/OligoWiz/) - which influences the noise and accuracy of analytical interpretation. The microarray probe sets and the size of probes including short ($15 \leq x < 30$ nucleotides), medium ($30 \leq x \leq 50$) and long ($x > 50$) oligonucleotides determine the flexibility, sensitivity, specificity and specimen coverage of the microarray

techniques. Short oligonucleotide probes are proper for microbial diagnostic microarrays (MDMs) (3,12,13).

Microarray spotter

The designed probes must be attached on solid surface of coated slides or other materials via spotter. The robotic system of spotter, spotting DNA fragments on coated slides with pins, ink jet printers or photolithography (1,7,11,14).

The spatial density of spotted DNA solution on coated slide, determines quantification, reproducibility, and sensitivity of microarray analytical outcomes (15).

For reducing the erroneous of data interpretation, there is a need for improvement in array processing details such as printing process of DNA probe spots to bind the oligonucleotides on the coated slides. Obviously, the level of data interpretation accuracy is directly in association with probe designing, probe set, probe length, chemistry of coated surface, printing methodology, solutions and washing protocols (1,13,15).

Any variation in DNA spots densities and concentrations, imbalances the hybridization process and results in noise and false data interpretation; because, each probe behaves as a DNA-meter and a natural sensor for its DNA target sequence within a sample (15-17).

Probe sets

To get a high qualified data and an accurate and unambiguous interpretation in microarray technology, there is a need for multi-dimensional improvements throughout the methodology. Among different parameters, the use of proper and updated probe set enables us to gain an appropriate data analysis. So, our efforts for improving probe sets give us a considerable ability to have a suitable outcome in association with microarray data analysis and interpretation (8,13,18).

There are many probe sets which can be used for detecting, identifying or mapping a determined target sequence. But sometimes, a common probe set may show an overlap more than 50% which is not acceptable in some cases. Today, there are several advanced databases including The International Nucleotide Sequence Database Collaboration (INSDC) which can support researchers for updating original probe sets to have an accurate, precise and qualified data analysis and interpretation (11).

In many cases, there are several probe sets for detecting a group of target sequences and vice versa. However, there are some cases that a probe set may act stronger and mapping unambiguous. To get a qualified and specific probe set, there are three criteria which must be seen via biologists during probe designing: specific probes for specific target sequences with strong alignment, high ability for detecting the highest number of splice isoforms, and detecting target sequences in the area of the 3' end of the nucleic acid strand. The probe set robustness score (S_r) which determines the sensitivity and the specificity of the probes and shows the probable erroneous of the raw data is calculated as below (13):

$$S_r = (1-p)^N$$

Where, p is the probable interruption of an individual base during the probe transcript synthesis and N, is the number of joint bases between probes of probe sets and labeled target sequences. The probe set overall score (S_o) can be calculated via the following equation (13):

$$S_o = S_s S_c S_r$$

There, S_s is the specificity score (the portion of probes pertaining to a probe set which detect specific target sequences), S_c is the coverage score (the ratio of labeled target sequences detected by probes relating to a probe set). Thus, a high and proper probe set overall score is directly correlated to high probe set robustness score, probe set specificity score and probe set coverage score.

There are different software tools such as jetset, that enable us to calculate the optimized level of probe

sets for detecting and identifying a specific target sequence. Simultaneously, the hybridization process is directly affected by probe sets and applied protocols (13).

The items that ensure the high efficacy of a probe set involve some biophysical properties, including probe calibration and noise calculation which are described below.

Probe calibration and noise calculation

Probe is a known sequenced DNA molecule which is immobilized on the solid surface of the microarray spotted slide. The unknown complementary DNA molecule free in the liquid phase solution, is labeled by fluorescent dyes like Cy3 and/or Cy5 and is identified as the target sequence (7,19).

If the distance of an immobilized probe from other contiguous probe molecules involves R_p region, so the R_p is obtained from the following equation:

$$R_p = \sqrt{\frac{1}{C_p \times 4}}$$

Here C_p is the number of combined DNA probes per unit on the solid surface of the slide. Chan et al. have reported a range of minimum to maximum values from $1R_p$ to 0.01 cm (19).

The size of immobilized DNA probe is directly related to the rate of target adsorption. With the increase of immobilized probe density per surface unit, the orbital efficacy reaches 1.0. This feature has direct effect on the quality and the quantity of hybridization. Besides, there is an ion currency around each immobilized probe on the chip which includes ions with two positions. The first ion group makes an extensive angular momentum with probes while the other group creates a narrow angular momentum with immobilized DNA molecules. Therefore, a sheath of ion must be seen in peripheral zone of the probes which affects the process of hybridization (19,20).

According to the Langmuir [1] and Freunlich [2] equations, it is possible to calibrate the designed probes in microarray technology (12):

$$[1] \quad y = \frac{y_{max}K_x}{1+K_x}$$

$$[2] \quad y = ax^b$$

Here, y_{max} is saturation level, K is binding constant, x is concentration, y is signal intensity and a, b are experimental parameters (12).

But, Pozhitkov et al. have shown the relative error including noise and uncertainty belonging to the mean signal intensity [*Err(y)*] in associated with probe calibration calculated via the Langmuir [1] and Freunlich [2] equations. Therefore, [*Err(y)*] is a key parameter to calculate the relative error pertaining to calculated concentration. Thus to have an accurate probe calibration, the error in associated with noise and uncertainty of probes must be calculated via the below equations (12):

$$Err(x) = \left(\frac{y_{max}}{y_{max}-y}\right)Err(y)$$

$$Err(x) = \frac{1}{b}Err(y)$$

Hybridization

DNA oilgonucleotides anchored to fragments of the solid surfaces via covalent bonds match to complementary targets in a specimen. The presence of complementary sequences may lead to form duplex strands of nucleic acids. Furthermore, the densities of surfaces influence the regional electrostatics and the efficacy of hybridization (15,21).

The physicochemical properties pertaining to probes, targets and chip surfaces enable researchers to predict the level of perfect-match in duplex forms of probe-target. The life-history of microarray, the type of array preprocessing and processing protocols, and the presence of some motifs based on GCCTCCC have direct effects on the quality of hybridizations and data interpretations by making noise (15).

Sometimes, the biophysical properties of DNA molecules may lead to noise and uncertainties in data interpretation. In accordance with previous studies,

balloon shaped loops may appear in microarray probes during the hybridization process. The occurred loops reduce the stability of duplex forms resulted from hybridization between of probes and targets (15). Moreover, an electrostatic repulsion is recognized between the DNA probes and targets which may interfere with the final outcomes and make errors in hybridization process and data interpretations (15,22). The repulsion force is predictable via the following equation (15):

$$E_r = wN_P Z_T \left(Z_P + \theta Z_T \right)$$

There, w refers to the electrostatic intermediated factor relating to NaCl density in the solution of hybridization, N_P is the probe surface intensity, Z_P is the number of bases pertaining to probe, Z_T is the number of bases pertaining to target and θ is the range of hybridization located between 0 and 1 (15).

Conflicts of interest: None declared.

Conclusion

Microarray technology is known as a revolutionary analytical molecular and genomic tool for large-scale applications in different fields such as medicine, molecular biology and genetics. Microarray fabrication and data interpretation are depending on several items such as probe set, bioinformatics approaches and biophysical properties. Seemingly, biophysicists, bioinformaticians and biologists must have intense collaboration with each-other to produce high-quality probe sets for decreasing microarray data analytical errors, noise and biases via qualified protocols and probe sets. Reducing noise, biases and uncertainty in different parts of microarray diagnostic methodology may lead to an increased reliability of this technology for progression of microarray applications in public laboratories and medical centers. An accurate, simple and unambiguous data analysis and interpretation is the final goal of molecular diagnostics in hospitals, medical care centers, and laboratories.

References

1. Behzadi P, Behzadi E, Ranjbar R. The application of Microarray in Medicine. ORL.ro 2014;24:24-6.

2. d'Hérouël AF. On diverse biophysical aspects of genetics. http://www.diva-portal.org/ smash/ get/ diva2:404329/ FULL-TEXT01.pdf (Accessed: January 23, 2015).

3. Kostic T, Butaye P, Schrenzel J (eds.). Detection of Highly Dangerous Pathogens: Microarray Methods for BSL 3 and BSL 4 Agents. John Wiley & Sons; 2009.

4. Behzadi P, Najafi A, Behzadi E, Ranjbar R. Detection and Identification of Clinical Pathogenic Fungi by DNA Microarray. Infectio.ro 2013;35:6-10.

5. Behzadi P, Ranjbar R, Alavian SM. Nucleic Acid-Based Approaches for Detection of Viral Hepatitis. Jundishapur J Microbiol 2015;8:e17449.

6. McLoughlin KS. Microarrays for pathogen detection and analysis. Brief Funct Genomics 2011;10:342-53.

7. Najafi A, Ram M, Ranjbar R. Macroarray: Principles & Applications. 1st ed. Tehran: Persian Science & Research Publisher; 2012.

8. Sandberg R, Larsson O. Improved precision and accuracy for microarrays using updated probe set definitions. BMC Bioinform 2007;8:48.

9. Arnandis-Chover T, Morais S, González-Martínez MÁ, Puchades R, Maquieira Á. High density MicroArrays on Blu-ray discs for massive screening. Biosens Bioelectron 2014;51:109-14.

10. Brun EM, Puchades R, Maquieira A. Gold, Carbon, and Aluminum Low-Reflectivity Compact Discs as Microassaying Platforms. Anal Chem 2013;85:4178-86.

11. Behzadi P, Behzadi E, Ranjbar R. Microarray Data Analysis. Alban Med J 2014;4:84-90.

12. Pozhitkov AE, Noble PA, Bryk J, Tautz D. A revised design for microarray experiments to account for experimental noise and uncertainty of probe response. PloS One 2014;9:e91295.

13. Li Q, Birkbak NJ, Gyorffy B, Szallasi Z, Eklund AC. Jetset: selecting the optimal microarray probe set to represent a gene. BMC bioinform 2011;12:474.

14. Relógio A, Schwager C, Richter A, Ansorge W, Valcárcel J. Optimization of oligonucleotide-based DNA microarrays. Nucleic Acids Res 2002;30:e51-e51.

15. Harrison A, Binder H, Buhot A, Burden CJ, Carlon E, Gibas C, et al. Physico-chemical foundations underpinning microarray and next-generation sequencing experiments. Nucleic Acids Res 2013:gks1358.

16. Lee CY, Harbers GM, Grainger DW, Gamble LJ, Castner DG. Fluorescence, XPS, and TOF-SIMS surface chemical state image analysis of DNA microarrays. J Am Chem Soc 2007;129:9429-38.

17. Eliot M. Getting the Noise Out of Gene Arrays. Science 2004;306:630-1.

18. Mieczkowski J, Tyburczy ME, Dabrowski M, Pokarowski P. Probe set filtering increases correlation between Affymetrix GeneChip and qRT-PCR expression measurements. BMC Bioinform 2010;11:104.

19. Chan V, Graves DJ, McKenzie SE. The biophysics of DNA hybridization with immobilized oligonucleotide probes. Biophys J 1995;69:2243-55.

20. Chen FF, Evans JD, Zawalski W. Calibration of Langmuir probes against microwaves and plasma oscillation probes. Plasma Sources Sci Technol 2012;21:055002.

21. Vainrub A, Pettitt BM. Thermodynamics of association to a molecule immobilized in an electric double layer. Chem Phys Lett 2000;323:160-6.

22. Vainrub A, Pettitt BM. Coulomb blockage of hybridization in two-dimensional DNA arrays. Phys Rev E 2002;66:041905.