

SURVEY AND SUMMARY

Specificity of DNA microarray hybridization: characterization, effectors and approaches for data correction

Hinanit Koltai^{1,*} and Carmiya Weingarten-Baror²

¹Department of Ornamental Horticulture, ARO Volcani Center, Bet Dagan and ²Department of Neurobiology, The Weizmann Institute of Science, Rehovot, Israel

Received December 23, 2007; Revised and Accepted February 11, 2008

ABSTRACT

Microarray-hybridization specificity is one of the main effectors of microarray result quality. In the present review, we suggest a definition for specificity that spans four hybridization levels, from the single probe to the microarray platform. For increased hybridization specificity, it is important to quantify the extent of the specificity at each of these levels, and correct the data accordingly. We outline possible effects of low hybridization specificity on the obtained results and list possible effectors of hybridization specificity. In addition, we discuss several studies in which theoretical approaches, empirical means or data filtration were used to identify specificity effectors, and increase the specificity of the hybridization results. However, these various approaches may not yet provide an ultimate solution; rather, further tool development is needed to enhance microarray-hybridization specificity.

INTRODUCTION

DNA microarray technology has evolved rapidly in the last 10 years. This technology enables conducting experiments that quantify gene expression on a large scale. As a result, functions have been assigned to previously unannotated genes and genes have been grouped into functional pathways (1,2). Nevertheless, a strict examination of this technology reveals that in some cases only limited data are produced.

For example, by comparing results from different microarray technologies, a significant difference was shown in the number of differentially expressed genes detected (3,4). Also, different platforms have been shown to possess large variance in reproducibility between

replicates, large variance in sensitivity to fluctuations in messenger copy number and variations in accuracy measured relative to real-time PCR (3,5). As a result, for example, multigenic disease classifiers that predict the clinical outcome of a particular disease, detected across microarray technologies, do not seem to be consistent (6–8). This inconsistency impairs the reliability of DNA microarrays as a diagnostic tool.

The discrepancies in microarray results are a consequence of differences in microarray measures, such as accuracy [i.e. ‘the degree of conformity of the measured quantity to its actual (true) value’; (6)], sensitivity [i.e. ‘the concentration range of target molecules in which accurate measurements can be made’; (6)], reproducibility [i.e. ‘the degree to which repeated measurements of the same quantity will show the same or similar results’; (6)] and specificity [i.e. ‘the ability of a probe to provide a signal that is influenced only by the presence of the target molecule’; (6)].

Accuracy, sensitivity and reproducibility may be affected by several effectors. These measures and their effectors are discussed by Dufva (9) and Draghici *et al.* (6), and will not be detailed here. An example for an effector of sensitivity, reproducibility and accuracy is the type of microarray platform (Box 1): oligonucleotide arrays have been found to be more reproducible and sensitive than cDNA arrays (3), and some oligonucleotide arrays have been found to be more accurate than others (5). Sensitivity is also affected by probe density (i.e. the number of different probes that are fabricated in a given area), which has been shown to be an effector for the availability of probes for hybridization (9,10); this availability may also be affected by the steric restrictions imposed by the solid microarray surface (11). A higher availability of probes for hybridization has been demonstrated to increase sensitivity (12). In addition, sensitivity is affected by the hybridization signal-to-noise ratio (i.e. the ratio between the spot signal and that of the background): a low background increases microarray hybridization sensitivity (13).

*To whom correspondence should be addressed. Tel: +972 3 9683039; Fax: +972 3 9669583; Email: hkoltai@agri.gov.il

Box 1. Microarray types

Most of the studies on hybridization specificity have been performed on two types of microarrays: short-oligomer microarrays (e.g. Affymetrix Genechips[®], NimbleGen[®], Agilent[®] and Febit[®]), and long-probe microarrays, i.e. cDNA microarrays, with probes of up to a few hundreds bases in length.

Short-oligomer microarrays may be either custom-designed or ready-made. Some of the array types are fabricated *in situ*, on the microarray slide (e.g. Affymetrix), whereas other types are spotted (e.g. Agilent). Probe design varies among microarrays. Some of the short-oligomer microarrays might contain oligomers arranged as probe-sets for the detection of target molecules of a single reference sequence, some of the probes within a probe-set are designed to detect a target molecule via its different segments. In some of the microarrays, spot-sets include also 'probe-pairs', (e.g. Affymetrix microarrays), each probe pair consists of perfect match (PM) and mismatch (MM) probes; the PM probe is designed to be complementary to a reference sequence; the MM probe is designed to be complementary to the reference sequence except for a one-base mismatch. Long-probe microarrays, like the cDNA microarrays, are usually designed such that one probe (cDNA clone) detects one complementary transcript. Usually, in cDNA microarrays, each target gene is represented by more than one cDNA spot.

The number of samples simultaneously hybridized to the microarrays might be either one or two; either absolute or comparative amounts of the target molecules are determined, respectively. For dual-color microarrays (e.g. NimbleGen, cDNA arrays), cohybridization is performed for two samples labeled with two different (fluorescent) colors. Hence, a scanned image of the microarray chip is generated for each color, and is used for comparative quantification of the target molecules. For Affymetrix arrays, quantification of target molecules is based on one color and produces absolute measures of target-molecule quantities.

Low specificity of microarray hybridizations has been suggested to be one of the prime measures affecting discrepancies in gene-expression profiles between different probes targeting the same region of a given transcript (6,14–17) or between different microarray platforms (18–20); in the present review, we will highlight the issue of microarray-hybridization specificity as a key measure that once improved, may increase the validity of microarray results.

Microarrays consist of multiple probes. Hence, a prime key for specificity during microarray hybridization, for either short-oligomer or cDNA microarrays (Box 1), is the ability of the probe to discriminate between different target molecules.

Probes are designed to be complementary to the target molecule according to the Watson–Crick rules of binding. Therefore, a probe with high specificity to its target molecule should provide a signal influenced only by the presence of the target molecule. Nevertheless, a perfect match in terms of sequence-similarity-based complementarity between a probe and its target molecule does not guarantee specificity. This is due to the presence of thousands of target molecules during microarray hybridization—each target molecule being composed of tens of hundreds or thousands of four-nucleotide bases, and to the effect of different effectors (discussed subsequently) of hybridization specificity, which may alter the ability of

a probe to bind to a target molecule. Hence, there is often some degree of microarray-probe hybridization to a target molecule which is not strictly complementary to it or *vice versa*, a variable number of target molecules that are hybridized to a microarray probe which is not exactly complementary to them.

It should be noted that the requirement of microarray-hybridization specificity may vary among studies, i.e. it may be dictated by the study application. Usually, a high level of hybridization specificity is required; for example, different transcripts are sought to discriminate between different members of a gene family for their quantification (15). An altered specificity requirement is given in the form of cross-species hybridization (CSH) studies (21), in which the target molecules and microarray probe are from different species. In these studies, on the one hand, there is a need to separate between two effectors of reduced hybridization specificity: gene expression and sequence differences (21). On the other, during CSH, nonspecific hybridization is required for the identification of cross-species genes [e.g. putative homologous or orthologous genes (22–26)].

In the present review, we define hybridization specificity across four levels of hybridization: the single probe, the single spot, the spot-set and the microarray platform. We then outline the possible effects of low hybridization specificity at each of these specificity levels on the obtained biological results. Next, effectors of hybridization specificity are discussed from two different vantage points: in the first, hybridization-specificity effectors defined in multiple studies are listed and their effects on specificity presented. In the second, different approaches to coping with hybridization specificity are discussed using example studies; these approaches include theoretical studies, empirical means or data filtration, aimed at defining and modeling effectors of hybridization specificity, and at increasing specificity of the hybridization results.

FOUR LEVELS OF HYBRIDIZATION SPECIFICITY

We define four levels of hybridization specificity in the context of microarray hybridization. The first is of hybridization between a single probe molecule and a single target molecule (Figure 1A). The two molecules may exhibit perfect hybridization (Figure 1Ai), partial hybridization (i.e. the target molecule is only partially hybridized to the probe; Figure 1Aii) or no hybridization (Figure 1Aiii).

The second level of specificity is of a spot (Figure 1B). At this level, multiple probe molecules that compose one spot are hybridized to multiple target molecules. The spot probes may exhibit perfect, partial or no hybridization with the target molecules (Figure 1Bi, Figure 1Bii and Figure 1Biii, respectively). Notably, at this level, partial hybridization may have one or both of two forms: only some of the probes may be hybridized to the target molecule, or probes may be hybridized to only some of the target molecules. This partial hybridization, at the spot level, may be a result of cross-hybridization (i.e. hybridization between sequences that are not strictly

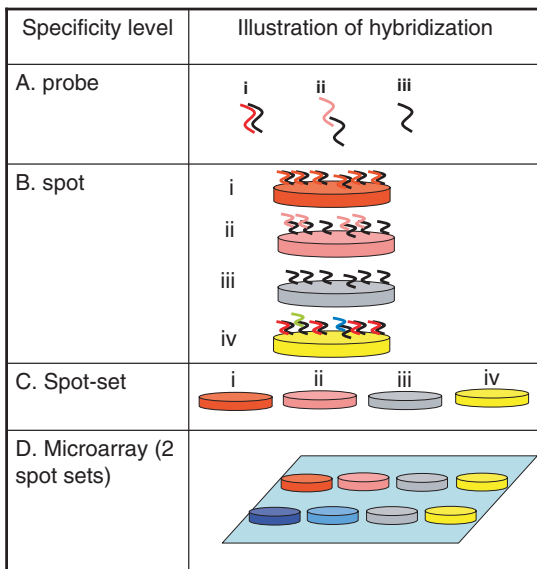


Figure 1. Illustration of the different levels of specificity of microarray hybridization. **(A)** Specificity at the probe level; matching between a single probe molecule and a single target molecule: perfect match (i), low match (ii), no match (iii). **(B)** Specificity at the spot level; matching between multiple probe molecules that make up one spot and multiple different (i.e. derived from more than one reference sequence) target molecules: perfect match (i, red spot), low match (ii, pink spot), no match (iii, gray spot) or cross-hybridization (iv, yellow spot). **(C)** Specificity at the spot-set level; matching between multiple spots that represent the same reference sequence, but through its different segments, and multiple different target molecules: perfect match (i), low match (ii), no match (iii) or cross-hybridization (iv). Black lines denote probe molecules of a single reference sequence that perfectly match the spot probes, pink lines denote target molecules of a single reference sequence with lower sequence similarity to the spot probes, green and blue lines denote target molecules of other reference sequences. **(D)** Specificity at the microarray level; matching between multiple spot-sets and multiple different target molecules: perfect match (red spot), low match (pink spot), no match (gray spot) or cross-hybridization (yellow spot).

complementary; Figure 1Biv), due to the presence and hybridization of nontarget molecules with sequences similar to that of the spot probes. Since a spot is composed of multiple probes, a single spot may simultaneously bear all combinations of one to four of the presented probe-target molecule types of binding.

The third level of specificity is of a spot-set [or, in Affymetrix terminology, ‘probe-set’ (Box 1); Figure 1C], in which multiple spots represent different segments of the same reference sequence (e.g. different exons of a gene). At this level, different spots of a spot-set may exhibit perfect hybridization with the target molecule (Figure 1Ci); partial hybridization with the target molecule (Figure 1Cii) due to the presence of probes with mismatches to the target molecule as a result of, for example, an annotation error in the gene sequence, or intended mismatches introduced to quantify nonspecific hybridization; no hybridization (Figure 1Ciii) due to, for example, alternative splicing of a transcript, leading to probes with no match to the target molecule; cross-hybridization (Figure 1Civ) due to, for example, a spot, within a spot-set that represents an evolutionarily

conserved gene segment, which hybridizes with nontarget molecules derived from various gene-family members.

The fourth level of specificity is that of the microarray, in which a variable number of spot-sets may exhibit different forms of hybridization with target sequences (Figure 1D): perfect hybridization (i.e. all target molecules are hybridized to their representative spot-sets and all spot-sets are hybridized to the target molecules they represent), partial hybridization in either direction, no hybridization (i.e. target molecules are not hybridized to any spot-set or spot-sets do not match any target molecules) or cross-hybridization (e.g. target molecules of different genes hybridize to the same spot-set or target molecules of a particular gene hybridize to several different genes’ spot-sets). These different forms may exist for a large number of different target molecules or spot-sets.

EFFECTS OF LOW HYBRIDIZATION SPECIFICITY ON THE OBTAINED BIOLOGICAL RESULTS

Low specificity at any of the four levels may affect the obtained biological results: an effect of low specificity at lower levels accumulates and influences the results at higher levels. For example, a single spot, which contains multiple probes, may simultaneously possess probes that are perfectly hybridized to the target sequence, probes that are partially hybridized, e.g. by cross-hybridization to nontarget molecules, and probes that are not hybridized to a target molecule. Hence, a combination of high- and low-specificity hybridizations may compose a spot signal, leading to only a poor reflection of the target molecule amount.

At the spot level, variation between spot hybridizations might lead to a condition in which some of the spots are perfectly hybridized with the target molecules and some are partially hybridized with the target molecules. Since a target molecule’s concentration at the spot-set level is determined as the average of the spot-set’s signal, results can be incorrect (i.e. lower than the true target molecule concentration; see, for example, reference 27). Once spots are hybridized by cross-hybridization, competition may occur: on the one hand, nontarget molecules that are hybridized by cross-hybridization to a spot-set may be less available to their target spot-set; on the other, they may block the cross-hybridized spot-set probes for specific hybridization with target molecules. As a result, spurious results for both involved spot-sets are expected.

At the microarray level, cumulative effects of partial hybridization, cross-hybridization and no hybridization may dramatically alter the genome-scale results obtained for an experiment. This has been evidenced by the differences in gene-expression profiles demonstrated between high hybridization specificity [of species-specific hybridization (SSH)] and low hybridization specificity (of CSH) results, for the same RNA sample (28).

Notably, the various levels and forms of hybridization specificity described above exist in a steady state, attained as a result of the effect of several effectors of microarray-hybridization specificity. These effectors are presented subsequently.

CATEGORIES AND EFFECTORS OF HYBRIDIZATION SPECIFICITY

Effectors of hybridization specificity have been previously recognized for nucleic acid hybridization, for a single or only a few probe identities [such as those used for Southern procedures;(29)]. Taking into account the multiple probes composing a microarray, a number of these probes are likely to hybridize by nonspecific hybridization; this leads to reduced specificity of microarray hybridization which does not exist for a single or only a few probe identities. Hence, multiple effectors of microarray-hybridization specificity should be considered; these are presented below.

Microarray-hybridization specificity can be divided into two categories with respect to associated effectors: the first includes the sequences of the probes and target molecules, and the second, the hybridization kinetics.

Sequences of probes and target molecules

Underlying the hybridization between a probe and a target molecule is the sequence complementarity between them. Effectors in this category include:

- (i) The level of matching between probe and target molecules (13,21,28,30)—here, although a clear-cut threshold may be difficult to define (28,31), in general, a higher level of matching between probe and target molecule enhances hybridization specificity.
- (ii) The length of the probe—short probes have higher specificity than longer ones (13).
- (iii) The G and C content—an increase in G and C content increases hybridization specificity (13,32).
- (iv) The primary molecular interactions that the probe may form with itself—increased thermodynamic stability of a probe's secondary structures with itself reduces hybridization specificity (33).
- (v) The length and composition of the contig formed during hybridization—Chen *et al.* (34) suggest that increased continuity of base pairs between probe and target sequences (i.e. pairing segments) increases specificity; He *et al.* (30) suggest that to achieve higher hybridization specificity, nonspecific hybridization stretches should be kept to a minimum.
- (vi) Neighboring bases along the probe and target molecule sequences—thermodynamic parameters for each two neighboring bases were established for DNA Watson–Crick pairing (summarized in reference 32). The nearest-neighbor modeling was shown to predict hybridization and melting for DNA and RNA sequences (35). Moreover, based on nearest-neighbor modeling, the binding free energy for a probe and target molecule was calculated, using established thermodynamic parameters, and was shown to be an effector of hybridization specificity (30).
- (vii) The position of matching (or mismatching) between probe and target molecules—the closing Watson–Crick pair on the mismatch, and its orientation has a large effect on hybridization. Some of the studies are summarized in reference 32: the nearest-neighbor

model may be extended to include the effect of neighboring base pairs (as thermodynamic parameters) for interactions between mismatches (32). In addition, internal mismatches may be thermodynamically either hybridization stabilizing or hybridization destabilizing, whereas all terminal mismatches are thermodynamically hybridization stabilizing (summarized in reference 32). Accordingly, mismatches at the center of the probe are more discriminating (i.e. have a stronger effect on specificity) than those at the ends of the probe (36). However, probes with segments complementary to segments at either end of the target molecule are suggested to have higher specificity than those that are complementary to segments within the central portion of the target molecule (13). He *et al.* (30) suggest that for higher specificity, i.e. to reduce nonspecific hybridization, the nonspecific hybridization stretches should be evenly distributed within the oligo probe.

Hybridization kinetics

The kinetics of hybridization depends on the quantity of target and nontarget molecules, the duration of hybridization and prehybridization and the input of energy, in the form of hybridization temperature. Also, hybridization solution mixing influences hybridization kinetics.

Quantity of target and nontarget molecules. Increasing sample complexity has been shown by Wick *et al.* (37) to lead to an increase in signal resulting from incorrect complexes of probes and nontarget molecules (i.e. hybridized by cross-hybridization), and to a decrease in signal resulting from correct hybridization complexes. Hence, an increase in the amount of nontarget molecules reduces specificity.

Duration of hybridization and prehybridization. Correct and incorrect spot hybridization have different hybridization kinetics: as a result of a mismatch to the probe, nontarget molecules are less strongly bound and therefore dissociate faster (36,38). Since a nontarget molecule dissociates from a hybridized probe faster than a target molecule, increased specificity is achieved with increased hybridization time (12,39,40). For the same reason, a short period of posthybridization washing, which allows dissociation of many of the incorrect complexes and only a few of the correct complexes, results in higher specificity (38).

Based on the kinetics of association and dissociation, two hybridization phases are defined by Bishop *et al.* (41). In each phase, there are certain effects of the competition between different target molecules hybridized to the same spot: during 'phase I' of hybridization, there is a reduction in binding sites due to hybridization of multiple targets; during 'phase II', there is displacement of lower affinity targets by higher affinity ones (41).

Input of energy in the form of hybridization temperature. The kinetics of association or dissociation of correct and incorrect probe-target molecule complexes

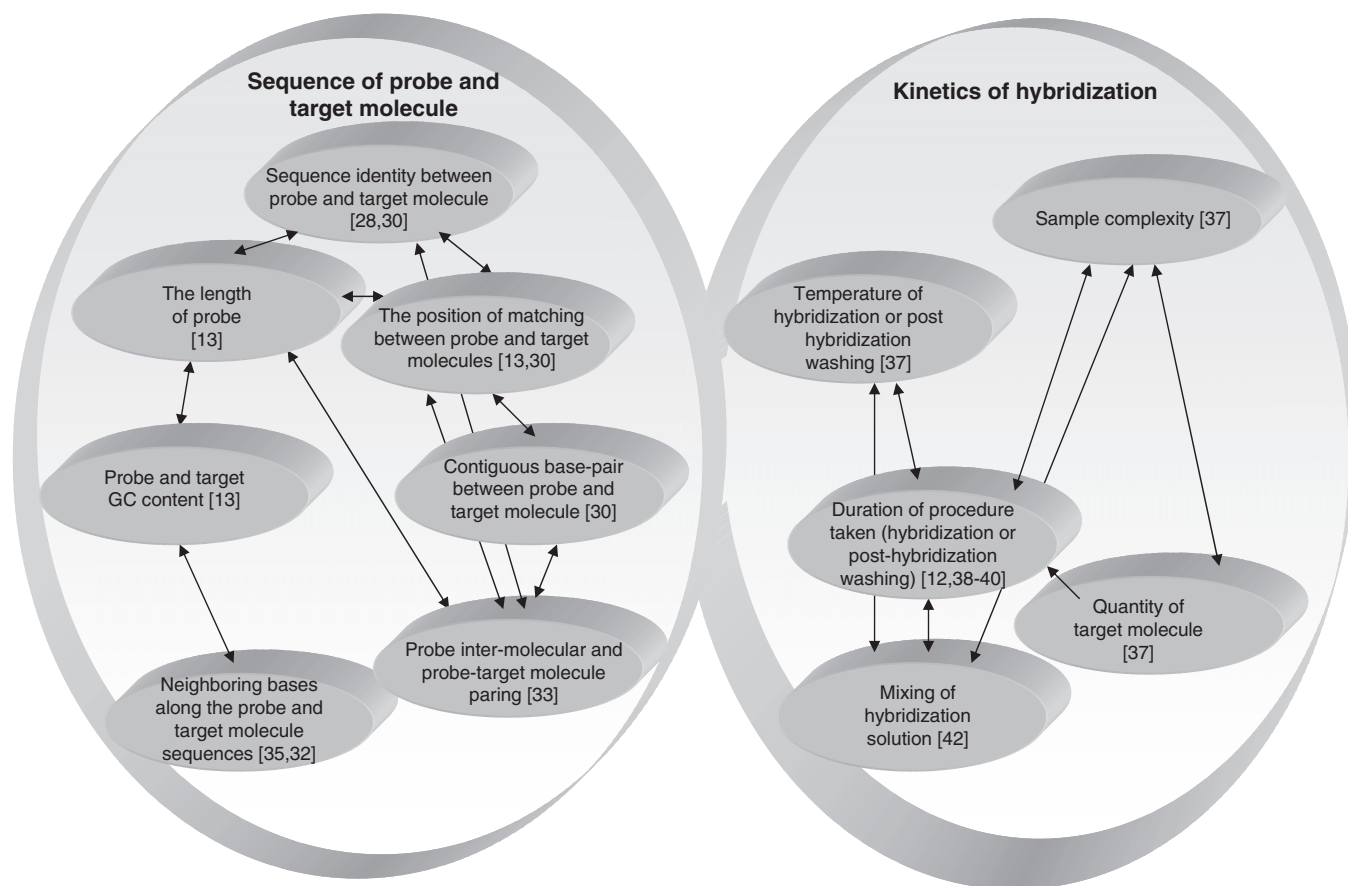


Figure 2. A schematic representation of various effectors of microarray-hybridization specificity, divided into categories, and their interconnections. Numbers in brackets refer to reference numbers. Associations between categories are provided in the text. Several examples for within-category interconnections (designated by arrows): within the category of 'kinetics of hybridization': temperature of hybridization or posthybridization washing affects the kinetics of complex generation, and the duration of hybridization or posthybridization needed for increased specificity. Within the category of 'sequence of probe and target molecule': the longer the probe, the more probe-target molecule pairing is expected. As a result, more matching in various positions is expected (13).

has been shown to be a function of the temperature of the hybridization or posthybridization washing (37).

Hybridization solution mixing. Mixing the hybridization solution during microarray hybridization leads to a higher specificity of hybridization (42,43). This conclusion was deduced from lower signal read from negative controls obtained by a mixing hybridization method in comparison to that obtained by a static hybridization method (42), and following examination of the effect of degree-of-mixing during hybridization on the discrimination between perfect and mismatch duplexes (43).

In summary, hybridization specificity is influenced by various specificity effectors; however, none of these effectors stands alone. Rather, specificity effectors are interconnected (Figure 2); effectors in the 'sequence of probe and target molecule' category are associated with the effectors of the 'hybridization kinetics' category. For example, sequence identity is associated with the quantity of target molecules: an increase in nontarget molecules reduces specificity (37). In addition, the kinetics is different for different matching between a probe and target molecule (38), whereas the kinetics of association or

dissociation of probe and target molecule is dependent on the quantity of target molecules (38).

Dimensions of specificity

For a thorough consideration of microarray-hybridization specificity, the above-described specificity effectors should be considered in association with the level of specificity and the procedural stages taken for microarray hybridization: probe design prehybridization, hybridization and posthybridization washing or data analysis.

Each of the specificity effectors may directly affect specificity at one of the specificity levels, and as a result, influence other specificity levels as well. For example: effectors of probe sequence, such as the position of match (or mismatch) between probe and target molecule, although determined at the probe level, may influence the signal, obtained at the spot level. This signal may be different between spot-sets, as each spot may contain probes with a certain position of matching, which differs between spots. Hence, matching position may influence the spot-set specificity level as well; differences implicated at the spot-set level may affect result specificity at the microarray level (discussed above).

Each of the effectors, at a certain specificity level, may reach a steady state, during one or more of the procedural stages taken for microarray hybridization. For example, the differences in hybridization kinetics of correct and incorrect complexes influence the specificity of the results during both hybridization and posthybridization washing (12,38,39,42,43). Another example is for effectors in the category of 'probe sequence' (e.g. sequence identity between probe and target molecule): these should be considered in both probe design prehybridization and probe filtration posthybridization.

Hence, all three dimensions, namely specificity effectors, levels of specificity and the procedural stages taken, comprise the space in which microarray-hybridization specificity should be considered.

STUDIES OF SPECIFICITY

With the aim of understanding and quantifying microarray-hybridization specificity, studies have either modeled various effectors or used empirical means to quantify them. Other studies have applied data filtration, based on some specificity effectors, for increased specificity of the obtained results. Examples of these studies are presented subsequently.

Theoretical studies of specificity

Modeling is a challenging task, designed to lead to an understanding of the phenomenon being modeled. A good model for hybridization specificity should, on the one hand, incorporate all of the hybridization-specificity effectors (outlined above) that affect the hybridization results, such that a prediction of target molecule concentration can be obtained up to inherent measurement noise (44–46). On the other hand, modeling should avoid over-parameterization (47), i.e. the number of parameters should be large enough to represent effectors of the study process, but small enough to allow generalizations.

Multiple studies have modeled specificity effectors (13,34,36,44,47–63). However, these studies include only some of the specificity effectors, and examine only a certain level of specificity. In the following, we outline several of these studies as examples of evaluations of hybridization specificity by modeling.

Modeling specificity at the probe level. As already detailed, this first level of specificity underlies all other specificity levels. Hence, its modeling may lead to a better understanding of hybridization specificity. Many of the studies modeling specificity at this level include some of the specificity effectors, in different combinations, for modeling and quantification. None, however, include all effectors, from all different categories, for specificity quantification. Three example studies are presented here, demonstrating the modeling of effectors from different categories.

The Carlon and Heim (48) model combined two specificity effectors: the position of matching and mismatching between probe and target molecule and the

molecular structures that may be formed during hybridization for a target molecule. They also included the stacking free energy of neighboring bases (all are effectors of the 'sequence of probe and target molecule' category). Their model had an advantage over others (50), in that it included the free-energy parameters of formation of RNA/DNA and RNA/RNA duplexes in solution, rather than DNA/DNA, for both perfect-match (PM) and mismatch (MM) probes (Box 1); hence, the model could potentially reflect, at least for expression arrays, hybridization between DNA probe and RNA target molecule or two RNA target molecules. Their model was able to quantify target molecule concentration.

Rather than considering two molecules (probe and target) in solution, Jayaraman *et al.*'s (13) study simulated the association between a single molecule attached to a hard surface and a single molecule in solution; the latter association could potentially represent microarray hybridization of a single probe and a single target molecule. Of the specificity effectors, they examined the length of the probe and the region of match or mismatch along the target and probe molecules (from the 'sequence of probe and target molecule' category), and were able to define effects of the examined effectors on hybridization specificity.

The above models took into account only the parameters of the association, during hybridization, between probe and target molecules. In contrast, Held *et al.* (44) included in their model the kinetics of association of a probe and a target molecule during hybridization and the kinetics of dissociation of a probe and a target molecule during posthybridization washing, for accurate predictions of gene-expression results. Their model was able to predict target molecule concentrations.

Modeling specificity at the spot level. The probe-level models are based on signals that are usually detected at the spot level. Hence, these two levels of specificity, namely probe and spot, may not be easily separated. Nevertheless, to the best of our knowledge, no model of hybridization specificity has yet been developed especially for the spot level.

Modeling specificity at the spot-set level. The specificity level of the probe-set comprises a well-defined level for the determination of hybridization specificity. This is because the presence of matched and mismatched spot pairs are designed (e.g. by Affymetrix) to confer a measurement of hybridization specificity, as demonstrated in the following example study.

Wu *et al.* (36) assessed the effect of the free-energy cost of hybridization of perfectly matched and mismatched probe pairs (from the 'sequence of probe and target molecule' category) on hybridization specificity. They showed that, for probe pairs, the energy cost of non-specific hybridization is much higher than that for specific hybridization. Consequently, they concluded that it would seem to be unreliable to use MM probe signals to track cross-hybridizing signals on a paired, PM probe.

Modeling specificity at the microarray level. Models at the microarray level of specificity consider effectors of hybridization specificity that are present once multiple probes are attached to a platform. This level of specificity is complex, since it includes effects of low specificity from all underlying specificity levels (discussed in the section 'Effects of low hybridization specificity on the obtained biological results'). Nevertheless, the extraction of biological results is usually based on this level of microarray specificity; hence, its improvement may have a practical effect on the quality of experimental conclusions.

The example presented here used Bayesian statistical modeling to examine the probability of multiple target molecules binding to a particular spot-set, for all spot-sets of the microarray, as a way of quantifying the initial amounts of RNA (54). These authors modeled hybridization specificity, excluding probe and target molecule sequence information, and examined, as effectors of hybridization, most of the procedural steps involved in the use of microarrays. Effectors included sample purity, array, pen, probe quantity, probe identification and probe length; only some of these (such as probe length) are effectors of specificity. Others (such as the pen used) are more likely to affect other measures of microarray hybridization, such as accuracy and reproducibility (detailed in the 'Introduction' section). A quantitative PCR supported their model (54).

In summary, modeling hybridization specificity by its effectors has led to a characterization of the effect of different specificity effectors, and in some cases, to predictions, to some extent, of target molecule amounts. However, Pozhitkov *et al.* (64) demonstrated only a partial predictive value for the modeling of various effectors of specificity, such as free energy of probe and target molecule hybridization and free energy of intramolecular folding: three nearest-neighbor models, modeling the various specificity effectors, only poorly explained the relationship between free-energy terms and signal-intensity values.

These discrepancies in the ability of the examined models to predict signals may be due to the complexity of microarray-hybridization specificity at its various levels (discussed earlier); perhaps only a coherent and complete model of hybridization specificity at all specificity levels, which will define all specificity effectors for each of the procedural stages taken for microarray hybridization, can fully predict signals with high specificity, and lead to the elimination of low specificity effects on hybridization results. However, the demonstrated ability of the models to represent, to various extents, the amount of target molecules by considering only some of the specificity effectors, may reflect associations between the specificity effectors (discussed above and in Figure 2). As a result of these associations, the need to include all specificity effectors may be reduced.

Empirical means for the detection of hybridization specificity

Until a coherent and complete model is created, empirical means are also needed to assess specificity. Although not

meant for generalization, these may be useful as practical means of assessing specificity on a case-by-case basis. In the following, we present some examples of studies which have measured hybridization specificity by empirical means. Some of the studies (27) use solely empirical means for specificity evaluation, while others combine theoretical studies with empirical approaches (30).

Empirical means for detecting hybridization specificity at the probe and spot levels. Empirical detection and quantification of probe hybridization is usually performed at the spot level. Hence, probe and spot levels may not be separable when empirical means are used for the quantification of target molecule amounts. Below are two example studies. Both combine the modeling of effectors at the probe level with empirical measurements of spots (30,31). Nevertheless, each study uses different specificity effectors and different spot measures.

He *et al.* (30) calculated the binding free-energy values of the probes, and correlated them to the relative signal intensity obtained for a spot. The results produced weighted parameters for each of the examined effectors (i.e. sequence identity, continuous stretch, free energy and mismatch position); these may be useful for a rational design of probes pre-microarray printing, or for filtration of microarray probe hybridization data post-microarray hybridization, for increased specificity.

Bar-Or *et al.*'s study (31) combined empirical detection of spot characteristics with a determination of the level of matching between probe and target molecule. These spot characteristics (e.g. spot-signal uniformity, spot dimensions), detected in an image of a scanned spotted cDNA microarray, were suggested to reflect the level of matching between probe and target molecules during CSH, establishing spot characteristics of image-quantification data as indicators of spot-level specificity during CSH.

Empirical means for detecting hybridization specificity at the spot-set level. The specificity level of the probe-set comprises a well-defined level for the determination of hybridization specificity not only because of the presence of matched and mismatched spot pairs (discussed above), but also because there are multiple spots within a spot-set, which are all designed to detect a single target molecule via its different segments. In this case, specificity may change between spots within a spot-set, as a function of, for example, degree of sequence conservation within a single gene.

For example, Hammond *et al.*'s (27) CSH study increased the specificity of the results by selecting for probe-sets (only perfectly matched probes were considered) with probes that possess increased sequence conservation between the reference and target species. Spot-sets with increased sequence conservation were empirically detected by genomic DNA (gDNA) hybridization of the target species to the microarray of the reference species, prior to the RNA CSH. Then, based on the gDNA hybridization efficiency, they selected for perfectly matched spot-sets, in which at least one probe hybridized to the target species DNA with an intensity that was above an arbitrary threshold. This enabled the identification of

all spot-sets that might hybridize to the target species transcripts with increased specificity, thereby providing an empirical assessment of the hybridization specificity of the examined CSH system.

Empirical means for detecting hybridization specificity at the microarray level. To cope with the complexity of specificity at the microarray level, theoretical studies at the probe level (to characterize specificity at its most basic level) have been combined with empirical studies at the microarray level of specificity (the level at which the results are usually extracted).

One such example is a study by Chen *et al.* (34), who hybridized one target sequence to multiple probes representing gene-family members. Microarray hybridization was empirically quantified at the microarray level, whereas 12 potential effectors of hybridization specificity at the probe level (such as probe length, probe and target GC content, percent sequence identity and overlap length between probe and target sequences) were modeled and their influence on the microarray results determined. Three multivariate statistical models (multiple linear regression, regression trees and artificial neural networks analysis) were used to quantify nonspecific hybridization. They found that the most contiguous base pairs between probe and target sequences (i.e. pairing segments) and the target GC content, more than percent sequence identity, were the prominent effectors of specificity.

Thus, both modeling and empirical approaches have demonstrated the ability to identify specificity effectors and quantify their effects on hybridization specificity. Nevertheless, for effective improvement of microarray data specificity, a rational filtration of the data, based on specificity effectors, which includes those originating from highly specific hybridization and excludes those originating from low- or nonspecific hybridization, should be applied. In the next section, we present some studies that filtered hybridization data to include only those with increased specificity, thus improving the specificity of the microarray results.

Filtration of hybridization data

In this section, we discuss the filtration of hybridization data which, based on filtration for specificity effectors, is aimed at increasing hybridization specificity. Here, the type of microarray (i.e. short-oligo or cDNA) is referred to in each of the studies; this is because for short oligoarrays (e.g. Affymetrix), filtration tools are particularly well developed (detailed below) relative to those for cDNA arrays. In addition, we indicate the basis for the filtration, i.e. whether it is based on results of modeling or quantification of specificity effectors, results of an empirical approach for the determination of hybridization specificity or results of homology searches.

Filtration of hybridization data at the spot level. For increased specificity, filtration based on a signal is better done at the spot-set or microarray levels (rather than at the single probe level), to facilitate comparative measurements of hybridization efficiency (i.e. hybridization may not be efficient for a single spot due to, for example, an air

bubble at that particular site). However, one study did filter hybridization data at the spot level. This was not done solely on a spot signal, but rather, on spot characteristics. In this study (31), spot characteristics of a cDNA array were demonstrated to be correlative to the level of matching between probe and target molecule (see details above). Hence, by filtering for spot characteristics, the filtering was for the level of matching, as an effector of specificity. This filtration produced improved clustering for two of the three examined experiments, involving SSH and CSH, enhancing the validity of the results. It was suggested that if a model were developed for specificity effects on spot characteristics, it could then lead to filtration of the data, with the ability to carefully control, via spot characteristics, the measure of hybridization specificity.

Filtration of hybridization data at the spot-set level. As already mentioned, the feature of a spot-set, which includes PM and MM probes, has been accompanied by the extensive development of tools for filtration-for-specificity (e.g. for Affymetrix microarrays), which is facilitated by the level of matching between probe and target molecule. Generally, three main normalization and filtration methods (RMA, dChip and MAS5) are used to correct for nonspecific hybridization of Affymetrix data (<http://www.affymetrix.com/>). For example, the RMA method is based on a statistical approach to the filtration of Affymetrix probe-sets: Irizarry *et al.* (55) suggested that mathematical subtraction of the MM probe intensity from the corresponding, paired PM probe intensity (as done in e.g. MAS5), does not translate to biological subtraction. Rather, use of only the PM values was advised, while accounting for the observed variation in intensity between the different spots within a probe-set.

Hammond *et al.* (27) filtered Affymetrix data based on an empirical approach. Based on the efficiency of gDNA hybridization (study detailed above), they selected PM probes of probe-sets in which at least one PM probe hybridized to the target species DNA with an intensity above an arbitrary threshold. CSH of *Brassica* RNA enabled filtering out data for PM probes that had low specificity, i.e. with gDNA hybridization levels below the chosen threshold.

Filtration of hybridization data at the microarray level. Filtration of hybridization data at the microarray level of specificity should rely on an effector that takes into consideration the hybridization specificity of all probes, spots and spot-sets (i.e. a microarray level effector). One such effector is the level of matching between probe and target molecule: although matching is absolutely determined, the cutoff used in filtering for increased specificity is determined relative to whole-microarray matching results. Two examples of such studies follow.

In one, the levels of matching between two different cDNA microarray platforms and between probe sequences and genomic data were determined, to identify probes that possess a high level of sequence similarity between reference microarray probes and target-species molecules, for CSH. This sequence-similarity-based

filtration led to improved matching of CSH and SSH transcriptomic results (28).

The second example of filtration for increased specificity at the microarray level, based on homology searches, was shown to be applicable to both cDNA arrays and oligoarrays. Flikka *et al.* (65) developed a tool for the detection of potential cross-hybridizations for DNA microarrays. The tool compares probe sequences with an extensive sequence database representing the transcriptome of the target species, followed by filtration of probes based on probe length and sequence-comparison results.

CONCLUDING REMARKS

In this review, we characterize microarray-hybridization specificity by defining its different levels, counting effectors of specificity and presenting several approaches, via example studies, aimed at improving hybridization specificity.

Nevertheless, several aspects of hybridization specificity may complicate its modeling and quantification and accordingly, data filtration. This complexity is made up of three dimensions (i.e. specificity levels, specificity effectors and the procedural stage at which a steady state is reached).

The measure of specificity, however, does not stand alone. Rather, other measures of microarray hybridization, namely accuracy, sensitivity and reproducibility, affect the obtained results. Moreover, these measures are associated with, and may be influenced by hybridization specificity: in the case of low hybridization specificity, the signal is subject to the presence of nontarget molecules, in addition to target molecules. The presence of nontarget molecules increases the complexity of the hybridization such that during a particular hybridization reaction, different molecules have different kinetics of hybridization. As a result, the results' accuracy [i.e. the degree of conformity of the measured quantity to its actual (true) value] may be lower. Once the accuracy is lower for a single hybridization reaction, the reproducibility between hybridization reactions, each with reduced accuracy, is expected to be lower. Association of lower specificity with lower reproducibility has been demonstrated by, e.g. Bar-Or *et al.* (31). Reduced specificity may also lead to reduced sensitivity: probe-binding affinity is reduced with reduced specificity, leading to reduced sensitivity (47). Also, sensitivity is affected by the solid microarray surface (10,11); including, in a model, a probe molecule tethered to a solid microarray surface enhanced modeling of microarray hybridization specificity (13).

To conclude, before using microarray hybridization experimentally, the issue of microarray-hybridization specificity should be thoroughly considered. First, a certain degree of specificity may be chosen in advance, before beginning the study. Second, to obtain this degree of specificity, specificity should be controlled. 'Controlled specificity' may be achieved by including many of the specificity effectors while designing the probes, and conducting hybridization and posthybridization reactions with filtering of the obtained results. Since specificity is

interconnected with the other microarray hybridization measures of accuracy, sensitivity and reproducibility, effectors of all measures should be considered for probe design and filtration. It might be that only effectors combination, each as a weighted parameter, will confer a desired specificity, accuracy, sensitivity and reproducibility for a particular study.

ACKNOWLEDGEMENTS

Funding to pay the Open Access publication charges for this article was provided by Israeli Ministry of Agriculture.

Conflict of interest statement. None declared.

REFERENCES

1. Stoughton, R.B. (2005) Applications of DNA microarrays in biology. *Annu. Rev. Biochem.*, **74**, 53–82.
2. Plomin, R. and Schalkwyk, L.C. (2007) Microarrays. *Dev. Sci.*, **10**, 19–23.
3. Yauk, C.L., Berndt, M.L., Williams, A. and Douglas, G.R. (2004) Comprehensive comparison of six microarray technologies. *Nucleic Acids Res.*, **32**, e124.
4. Tan, P.K., Downey, T.J., Spitznagel, E.L., Jr, Xu, P., Fu, D., Dimitrov, D.S., Lempicki, R.A., Raaka, B.M. and Cam, M.C. (2003) Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res.*, **31**, 5676–5684.
5. Shippy, R., Sendera, T.J., Lockner, R., Palaniappan, C., Kaysser-Kranich, T., Watts, G. and Alsobrook, J. (2004) Performance evaluation of commercial short-oligonucleotide microarrays and the impact of noise in making cross-platform correlations. *BMC Genomics*, **5**, 61.
6. Draghici, S., Khatri, P., Eklund, A.C. and Szallasi, Z. (2006) Reliability and reproducibility issues in DNA microarray measurements. *Trends Genet.*, **22**, 101–109.
7. Wang, Y., Klijn, J.G., Zhang, Y., Sieuwerts, A.M., Look, M.P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M.E., Yu, J. *et al.* (2005) Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet*, **365**, 671–679.
8. Naderi, A., Teschendorff, A.E., Barbosa-Morais, N.L., Pinder, S.E., Green, A.R., Powe, D.G., Robertson, J.F., Aparicio, S., Ellis, I.O. *et al.* (2007) A gene-expression signature to predict survival in breast cancer across independent data sets. *Oncogene*, **26**, 1507–1516.
9. Dufva, M. (2005) Fabrication of high quality microarrays. *Biomol. Eng.*, **22**, 173–184.
10. Peterson, A.W., Heaton, R.J. and Georgiadis, R. (2001) The effect of surface probe density on DNA hybridization. *Nucleic Acids Res.*, **29**, 5163–5168.
11. Southern, E., Mir, K. and Shchepinov, M. (1999) Molecular interactions on microarrays. *Nature Genet.*, **21**, 5–9.
12. Dorris, D.R., Nguen, A., Gieser, L., Lockner, R., Lublinsky, A., Patterson, M., Touma, E., Sendera, T.J., Elghanian, R. and Mazumder, A. (2003) Oligodeoxyribonucleotide probe accessibility on a three-dimensional DNA microarray surface and the effect of hybridization time on the accuracy of expression ratios. *BMC Biotechnol.*, **3**, 6.
13. Jayaraman, A., Hall, C.K. and Genzer, J. (2006) Computer simulation study of molecular recognition in model DNA microarrays. *Biophys. J.*, **91**, 2227–2236.
14. Zhang, J., Finney, R.P., Clifford, R.J., Derr, L.K. and Buetow, K.H. (2005) Detecting false expression signals in high-density oligonucleotide arrays by an in silico approach. *Genomics*, **85**, 297–308.
15. Xu, W., Bak, S., Decker, A., Paquette, S.M., Feyereisen, R. and Galbraith, D.W. (2001) Microarray-based analysis of gene expression in very large gene families: the cytochrome P450 gene superfamily of *Arabidopsis thaliana*. *Gene*, **272**, 61–74.

16. Evertsz, E.M., Au-Young, J., Ruvolo, M.V., Lim, A.C. and Reynolds, M.A. (2001) Hybridization cross-reactivity within homologous gene families on glass cDNA microarrays. *Biotechniques*, **31**, 1182, 1184, 1186 passim.
17. Miller, N.A., Gong, Q., Bryan, R., Ruvolo, M., Turner, L.A. and LaBrie, S.T. (2002) Cross-hybridization of closely related genes on high-density microarrays. *Biotechniques*, **32**, 620–625.
18. Mecham, B.H., Klus, G.T., Strovel, J., Augustus, M., Byrne, D., Bozso, P., Wetmore, D.Z., Mariani, T.J., Kohane, I.S. and Szallasi, Z. (2004) Sequence-matched probes produce increased cross-platform consistency and more reproducible biological results in microarray-based gene expression measurements. *Nucleic Acids Res.*, **32**, e74.
19. Stafford, P. and Brun, M. (2007) Three methods for optimization of cross-laboratory and cross-platform microarray expression data. *Nucleic Acids Res.*, **35**, e72.
20. Cheadle, C., Becker, K.G., Cho-Chung, Y.S., Nesterova, M., Watkins, T., Wood, W., III, Prabhu, V. and Barnes, K.C. (2007) A rapid method for microarray cross platform comparisons using gene expression signatures. *Mol. Cell. Probes*, **21**, 35–46.
21. Bar-Or, C., Czosnek, H. and Koltai, H. (2007) Cross-species microarray hybridizations: a developing tool for studying species diversity. *Trends Genet.*, **23**, 200–207.
22. Brodsky, L.I., Jacob-Hirsch, J., Avivi, A., Trakhtenbrot, L., Zeligson, S., Amarglio, N., Paz, A., Korol, A.B., Band, M., Rechavi, G. et al. (2005) Evolutionary regulation of the blind subterranean mole rat, *Spalax*, revealed by genome-wide gene expression. *Proc. Natl Acad. Sci. USA*, **102**, 17047–17052.
23. Rifkin, S.A., Kim, J. and White, K.P. (2003) Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat. Genet.*, **33**, 138–144.
24. Rise, M.L., von Schalburg, K.R., Brown, G.D., Mawer, M.A., Devlin, R.H., Kuipers, N., Busby, M., Beetz-Sargent, M., Alberto, R., Gibbs, A.R. et al. (2004) Development and application of a salmonid EST database and cDNA microarray: data mining and interspecific hybridization characteristics. *Genome Res.*, **14**, 478–490.
25. Adjaye, J., Herwig, R., Herrmann, D., Wruck, W., Benkahl, A., Brink, T.C., Nowak, M., Carnwath, J.W., Hultschig, C., Niemann, H. et al. (2004) Cross-species hybridisation of human and bovine orthologous genes on high density cDNA microarrays. *BMC Genomics*, **5**, 83.
26. Fang, H., Tong, W., Perkins, R., Shi, L., Hong, H., Cao, X., Xie, Q., Yim, S.H., Ward, J.M., Pitot, H.C. et al. (2005) Bioinformatics approaches for cross-species liver cancer analysis based on microarray gene expression profiling. *BMC Bioinformatics*, **6** (Suppl 2), S6.
27. Hammond, J.P., Broadley, M.R., Craigon, D.J., Higgins, J., Emmerson, Z.F., Townsend, H.J., White, P.J. and May, S.T. (2005) Using genomic DNA-based probe-selection to improve the sensitivity of high-density oligonucleotide arrays when applied to heterologous species. *Plant Methods*, **1**, 10.
28. Bar-Or, C., Bar-Eyal, M., Gal, T.Z., Kapulnik, Y., Czosnek, H. and Koltai, H. (2006) Derivation of species-specific hybridization-like knowledge out of cross-species hybridization results. *BMC Genomics*, **7**, 110.
29. Wolcott, M.J. (1992) Advances in nucleic acid-based detection methods. *Clin. Microbiol. Rev.*, **5**, 370–386.
30. He, Z., Wu, L., Li, X., Fields, M.W. and Zhou, J. (2005) Empirical establishment of oligonucleotide probe design criteria. *Appl. Environ. Microbiol.*, **71**, 3753–3760.
31. Bar-Or, C., Novikov, E., Reiner, A., Czosnek, H. and Koltai, H. (2007) Utilizing microarray spot characteristics to improve cross-species hybridization results. *Genomics*, **90**, 636–645.
32. SantaLucia, J. and Hicks, D. (2004) The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 415–440.
33. Matveeva, O.V., Shabalina, S.A., Nemtsov, V.A., Tsodikov, A.D., Gesteland, R.F. and Atkins, J.F. (2003) Thermodynamic calculations and statistical correlations for oligo-probes design. *Nucleic Acids Res.*, **31**, 4211–4217.
34. Chen, Y.A., Chou, C.C., Lu, X., Slate, E.H., Peck, K., Xu, W., Voit, E.O. and Almeida, J.S. (2006) A multivariate prediction model for microarray cross-hybridization. *BMC Bioinformatics*, **7**, 101.
35. Dimitrov, R.A. and Zuker, M. (2004) Prediction of hybridization and melting for double-stranded nucleic acids. *Biophys. J.*, **87**, 215–226.
36. Wu, C., Carta, R. and Zhang, L. (2005) Sequence dependence of cross-hybridization on short oligo microarrays. *Nucleic Acids Res.*, **33**, e84.
37. Wick, L.M., Rouillard, J.M., Whittam, T.S., Gulari, E., Tiedje, J.M. and Hashsham, S.A. (2006) On-chip non-equilibrium dissociation curves and dissociation rate constants as methods to assess specificity of oligonucleotide probes. *Nucleic Acids Res.*, **34**, e26.
38. Zhang, Y., Hammer, D.A. and Graves, D.J. (2005) Competitive hybridization kinetics reveals unexpected behavior patterns. *Biophys. J.*, **89**, 2950–2959.
39. Dai, H., Meyer, M., Stepaniants, S., Ziman, M. and Stoughton, R. (2002) Use of hybridization kinetics for differentiating specific from non-specific binding to oligonucleotide microarrays. *Nucleic Acids Res.*, **30**, e86.
40. Sorokin, N.V., Chechetkin, V.R., Livshits, M.A., Pan'kov, S.V., Donnikov, M.Y., Gryadunov, D.A., Lapa, S.A. and Zasedatelev, A.S. (2005) Discrimination between perfect and mismatched duplexes with oligonucleotide gel microchips: role of thermodynamic and kinetic effects during hybridization. *J. Biomol. Struct. Dyn.*, **22**, 725–734.
41. Bishop, J., Wilson, C., Chagovetz, A.M. and Blair, S. (2007) Competitive displacement of DNA during surface hybridization. *Biophys. J.*, **92**, L10–L12.
42. Schupp, C.J., Jiang, G., Myers, T.G. and Wilson, M.A. (2005) Active mixing during hybridization improves the accuracy and reproducibility of microarray results. *Biotechniques*, **38**, 117–119.
43. Sorokin, N.V., Yurasov, D.Y., Cherepanov, A.I., Kozhekbaeva, J.M., Chechetkin, V.R., Gra, O.A., Livshits, M.A., Nasedkina, T.V. and Zasedatelev, A.S. (2007) Effects of external transport on discrimination between perfect and mismatch duplexes on gel-based oligonucleotide microchips. *J. Biomol. Struct. Dyn.*, **24**, 571–578.
44. Held, G.A., Grinstein, G. and Tu, Y. (2006) Relationship between gene expression and observed intensities in DNA microarrays—a modeling study. *Nucleic Acids Res.*, **34**, e70.
45. Naef, F. and Magnasco, M.O. (2003) Solving the riddle of the bright mismatches: labeling and effective binding in oligonucleotide arrays. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.*, **68**, 011906.
46. Tu, Y., Stolovitzky, G. and Klein, U. (2002) Quantitative noise analysis for gene expression microarray experiments. *Proc. Natl Acad. Sci. USA*, **99**, 14031–14036.
47. Zhang, L., Miles, M.F. and Aldape, K.D. (2003) A model of molecular interactions on short oligonucleotide microarrays. *Nat. Biotechnol.*, **21**, 818–821.
48. Carlon, E. and Heim, T. (2006) Thermodynamics of RNA/DNA hybridization in high-density oligonucleotide microarrays. *Physica A*, **362**, 433–449.
49. Hekstra, D., Taussig, A.R., Magnasco, M. and Naef, F. (2003) Absolute mRNA concentrations from sequence-specific calibration of oligonucleotide arrays. *Nucleic Acids Res.*, **31**, 1962–1968.
50. Held, G.A., Grinstein, G. and Tu, Y. (2003) Modeling of DNA microarray data by using physical properties of hybridization. *Proc. Natl Acad. Sci. USA*, **100**, 7575–7580.
51. Levicky, R. and Horgan, A. (2005) Physicochemical perspectives on DNA microarray and biosensor technologies. *Trends Biotechnol.*, **23**, 143–149.
52. Mei, R., Hubbell, E., Bekiranov, S., Mittmann, M., Christians, F.C., Shen, M.M., Lu, G., Fang, J., Liu, W.M., Ryder, T. et al. (2003) Probe selection for high-density oligonucleotide arrays. *Proc. Natl Acad. Sci. USA*, **100**, 11237–11242.
53. Binder, H., Kirsten, T., Loeffler, M. and Stadler, P.F. (2004) Sensitivity of microarray oligonucleotide probes: variability and effect of base composition. *J. Phys. Chem. B*, **108**, 18003–18014.
54. Frigessi, A., van de Wiel, M.A., Holden, M., Svendsrud, D.H., Glad, I.K. and Lyng, H. (2005) Genome-wide estimation of transcript concentrations from spotted cDNA microarray data. *Nucleic Acids Res.*, **33**, e143.
55. Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B. and Speed, T.P. (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.*, **31**, e15.
56. Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U. and Speed, T.P. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4**, 249–264.

57. Li,C. and Wong,W.H. (2001) Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl Acad. Sci. USA*, **98**, 31–36.
58. Binder,H. and Preibisch,S. (2005) Specific and nonspecific hybridization of oligonucleotide probes on microarrays. *Biophys. J.*, **89**, 337–352.
59. Zhou,Y. and Abagyan,R. (2002) Match-only integral distribution (MOID) algorithm for high-density oligonucleotide array analysis. *BMC Bioinformatics*, **3**, 3.
60. Lazaridis,E.N., Sinibaldi,D., Bloom,G., Mane,S. and Jove,R. (2002) A simple method to improve probe set estimates from oligonucleotide arrays. *Math. Biosci.*, **176**, 53–58.
61. Lemon,W.J., Palatini,J.J., Krahe,R. and Wright,F.A. (2002) Theoretical and experimental comparisons of gene expression indexes for oligonucleotide arrays. *Bioinformatics*, **18**, 1470–1476.
62. Chu,T.M., Weir,B. and Wolfinger,R. (2002) A systematic statistical linear modeling approach to oligonucleotide array experiments. *Math. Biosci.*, **176**, 35–51.
63. Chudin,E., Walker,R., Kosaka,A., Wu,S.X., Rabert,D., Chang,T.K. and Kreder,D.E. (2002) Assessment of the relationship between signal intensities and transcript concentration for Affymetrix GeneChip arrays. *Genome Biol.*, **3**, RESEARCH0005.
64. Pozhitkov,A., Noble,P.A., Domazet-Loso,T., Nolte,A.W., Sonnenberg,R., Staehler,P., Beier,M. and Tautz,D. (2006) Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. *Nucleic Acids Res.*, **34**, e66.
65. Flikka,K., Yadetie,F., Laegreid,A. and Jonassen,I. (2004) XHM: a system for detection of potential cross hybridizations in DNA microarrays. *BMC Bioinformatics*, **5**, 117.