

Method

Utilizing microarray spot characteristics to improve cross-species hybridization results

Carmiya Bar-Or^{a,b}, Eugene Novikov^c, Anat Reiner^d, Henryk Czosnek^b, Hinanit Koltai^{a,*}

^a Department of Ornamental Horticulture, ARO Volcani Center, Bet Dagan 50250, Israel

^b The Robert H. Smith Institute of Plant Sciences and Genetics in Agriculture, Faculty of Agricultural, Food and Environmental Quality Sciences, The Hebrew University of Jerusalem, Rehovot, Israel

^c Service Bioinformatique, Institut Curie, Paris, France

^d Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot, Israel

Received 7 March 2007; accepted 21 June 2007

Available online 20 September 2007

Abstract

Cross-species hybridization (CSH), i.e., the hybridization of a (target) species RNA to a DNA microarray that represents another (reference) species, is often used to study species diversity. However, filtration of CSH data has to be applied to extract valid information. We present a novel approach to filtering the CSH data, which utilizes spot characteristics (SCs) of image-quantification data from scanned spotted cDNA microarrays. Five SCs that were affected by sequence similarity between probe and target sequences were identified (designated as BS-SCs). Filtration by all five BS-SC thresholds demonstrated improved clustering for two of the three examined experiments, suggesting that BS-SCs may serve for filtration of data obtained by CSH, to improve the validity of the results. This CSH data-filtration approach could become a promising tool for studying a variety of species, especially when no genomic information is available for the target species.

© 2007 Elsevier Inc. All rights reserved.

Keywords: cDNA microarrays; Gene expression; Image quantification; Species diversity; Transcriptome

Cross-species hybridization (CSH), i.e., the hybridization of a (target) species RNA to a DNA microarray that represents another (reference) species, is widely used in comparative ecological and evolutionary studies, as well as when no representative microarray platform is available for the target species [1]. However, microarrays are designed for species-specific hybridizations (SSHs; [2]), not CSHs. Hence, CSH is still considered a nonstandard application of microarrays, and its results need to be carefully interpreted. Indeed, previous studies have produced contradictory results regarding the ability of CSH to reflect valid biological results (e.g., [3,4]). Rather, filtration of CSH data, to gain SSH-like data, has to be applied to extract valid information [1].

During CSH, a low level of sequence matching between transcripts of the target species and some of the microarray probes is expected [1]. In cases in which sufficient genomic information is available for the target species, its proper

utilization for the filtration of data from CSH to spotted cDNA microarrays has been shown to improve the results [5,6]. However, for cases in which genomic information is lacking for a target species (most species), no data-filtration approach for CSH to spotted cDNA microarrays is available.

cDNA microarray spots are composed of multiple probe molecules; during CSH, many of the probes are expected to have a low level of matching to the transcripts of the target species. We therefore hypothesized that the low-level probe–transcript matching during CSH will have cumulative effects on hybridized spots. These cumulative effects might be evident for a spot detected in an image of a scanned spotted cDNA microarray in the form of spot characteristics (SCs; e.g., spot signal uniformity or spot dimensions); SC values would differ between spots bearing probes with high and low probe–transcript matching. If this hypothesis is correct, then detection of SCs for a given spot could indicate the level of matching between a probe and the target transcript. Data filtration that includes those corresponding to spots with high probe–transcript match could then be performed by SC values. Such

* Corresponding author. Fax: +972 3 9669583.

E-mail address: hkoltai@agri.gov.il (H. Koltai).

filtration might lead to the extraction of valid data from CSH. Filtration by spot dimensions (i.e., spot diameter) has been performed for SSH [7].

Here, we present several lines of evidence in support of our hypothesis and, as a consequence, a novel data-filtration approach that utilizes SCs to improve CSH results. Since only SCs (and no genomic data) are needed for the data filtration, this approach could become a powerful tool for the extraction of valid data from CSH for a variety of species.

Results

Identification of bit-score-correlated SCs

To test our hypothesis of cumulative effects of low-level probe–transcript matching on a microarray spot, and the suggested data-filtration approach, SCs affected by sequence similarity between probe and target sequences were sought.

In a previous study [5], the level of sequence similarity between probes and transcripts was proven to be key to the derivation of SSH-like knowledge (and hence valid biological knowledge) from CSH data. The level of sequence similarity was determined between reference-species microarray probes and target-species genes: for each cDNA microarray clone, the sequence of the best clone-representative tentative consensus (TC) of the reference species was matched with the target species TCs, using the Basic Local Alignment Search Tool (BLAST). The sequence-similarity level, in terms of BLAST bit-score values, was used to filter the CSH data for those corresponding to probes with a high match to target transcripts, thereby facilitating extraction of SSH-like knowledge from CSH [5].

Therefore, in our search for SCs that are affected by sequence similarity between probe and target sequences, we looked for BLAST bit-score-correlated SCs (BS-SCs), i.e., SCs whose values correlate with high or low values of probe–transcript match: matching was detected by BLAST between TCs of the reference and of the target species and was scored by BLAST bit-score values. These sequence-similarity-affected SCs were then examined as parameters for data filtration that corresponds to probes with a high match to target transcripts.

To identify BS-SCs, publicly available TIGR (The Institute for Genomic Research) scanned microarray images from three experiments, designated 054, 057, and 058 (<http://www.tigr.org/tdb/potato/>), were examined. These experiments involved time-point profiling of gene expression in plants responding to salt, heat, and cold stress, respectively. In each experiment, RNA from seven Solanaceae plants (potato, tomato, eggplant, pepper, tobacco, *Nicotiana benthamiana*, and petunia) was subjected to SSH and CSH to TIGR potato spotted cDNA microarrays. MAIA software [8] was chosen for quantification of the microarray images, since it provides values for 10 different SCs.

Sufficient genomic information was available for only two (tomato and potato) of the seven Solanaceae species (<http://compbio.dfci.harvard.edu/tgi/plant.html>). Therefore, bit-score values were determined for sequence matching only between tomato TCs and microarray-represented potato TCs.

Notably, SC analyses were performed on potato clones, represented by the microarray spots. However, TCs were used for bit-score value determinations (see above). Consequently, the BS-SC results refer to both clones and TCs. For simplicity, both clones and TCs will be referred to as “genes.”

To find BS-SCs, values of the 10 SCs were determined [8] for each spot of the tomato-potato CSH (i.e., tomato RNA hybridized to potato microarray) for each of the three experiments. Each of the obtained SC values was plotted against BLAST bit-score values, the latter obtained for each spot as the matching value for matching between the spot-represented potato TC and the best BLAST-matched tomato TC.

Five of the 10 SCs were found to be correlated to bit scores (BS-SCs); these were found across all three examined experiments for the tomato-potato CSH data (Fig. 1; Supplementary Data 1), although the correlation to the bit score was incomplete. Four of the BS-SCs were measures of a single spot. These included Det—coefficient of determination between the intensities of the Cy3 and the Cy5 channels, Dia—spot diameter, GSym—spot geometrical symmetry, and CVR—coefficient of variation of two gene-expression ratio estimates: one by a linear regression approach and the other by a segmentation algorithm [8]. The fifth BS-SC was a coefficient of variation (CV) of gene-expression ratios obtained by replicated spots.

To test our hypothesis further, we sought to examine the correlation between the identified BS-SCs and the CSH data of a species that is more phylogenetically distant from potato than tomato. Hence, BS-SC values of a petunia-potato CSH were plotted against the bit-score values obtained by the tomato-potato TC sequence comparison. Among the examined species, petunia is considered the most phylogenetically distant from potato (the reference species) ([9]; http://www.sgn.cornell.edu/about/about_solanaceae.pl). Hence, for the petunia-potato CSH data, we expected a lower correlation between the five BS-SC values and the tomato-potato bit-score values, relative to that obtained for the tomato-potato CSH data. Indeed, a reduced correlation was observed (Fig. 1), suggesting that the BS-SCs reflect a CSH effect (i.e., effect of matching between multiple probes and transcripts on a transcriptomics scale; [1]), which is related to phylogenetic distance. Some degree of correlation still existed between the five BS-SC values and the petunia-potato CSH data due to a degree of sequence similarity between petunia and tomato genomes (both being Solanaceae family members [9]).

Taken together, the results obtained for the tomato-potato and petunia-potato CSH data strengthened the concept of an effect of gene-sequence similarity between reference and target species on the five BS-SCs. The question then became, can BS-SCs, similar to bit-score values, be used as parameters for the filtration of CSH data to improve result validity?

Determination of BS-SC thresholds

Thresholds were determined for the BS-SC-based data filtration. Since bit-score values were incompletely correlated to the BS-SCs, a determination of BS-SC thresholds based on a particular bit-score value [5] was impractical. Hence, on one

Download English Version:

<https://daneshyari.com/en/article/2821487>

Download Persian Version:

<https://daneshyari.com/article/2821487>

[Daneshyari.com](https://daneshyari.com)