Genes Nutr (2009) 4:123–127
DOI 10.1007/s12263-009-0115-8

BRIEF COMMUNICATION

An integrated bioinformatics approach to improve two-color microarray quality-control: impact on biological conclusions

Rachel I. M. van Haaften · Cristina Luceri · Arie van Erk · Chris T. A. Evelo

Received: 10 February 2009 / Accepted: 13 February 2009 / Published online: 10 March 2009

Abstract Omics technology used for large-scale measurements of gene expression is rapidly evolving. This work pointed out the need of an extensive bioinformatics analyses for array quality assessment before and after gene expression clustering and pathway analysis. A study focused on the effect of red wine polyphenols on rat colon mucosa was used to test the impact of quality control and normalisation steps on the biological conclusions. The integration of data visualization, pathway analysis and clustering revealed an artifact problem that was solved with an adapted normalisation. We propose a possible point to point standard analysis procedure, based on a combination of clustering and data visualization for the analysis of microarray data.

Keywords Microarray · Polyphenols · Normalisation · Quality control

Introduction

Rapid evolution occurs for microarray technology, used for large-scale measurements of gene expression at mRNA level in biomedical research. Studies using this technology yield huge amounts of data which have to be analyzed in a correct way to eventually give useful information about the physiological outcome of the experiment. In the process from array production to final physiological outcome of a microarray experiment, numerous things can have a large impact on the interpretation of the final results of an experiment.

The construction of a microarray requires the production of a large number of correct probes and accurate spotting of the probes onto the glass slides. Many factors can influence the spotting, e.g., blocked spotting pins, glass slide surface treatment and environmental conditions [1–3].

Those and other technical issues during microarray preparation can influence the spot quality which can be detected after image analysis of the scanned microarray images. Spot quality can be documented by, e.g., signal-to-noise ratio, spot size irregularity, intensity saturation status, intensity distribution issues as a consequence of non-specific binding or irregular distribution of the printed DNA on the slide, morphological issues and background issues [4, 5]. Next to the production of the microarray the final results of an experiment can also be influenced by the quality of the initial RNA sample before hybridization and by the researcher performing the actual hybridization of the sample onto the array [6–8]. Some of the sources of variation can be removed or minimized by removing bad spots from further analyses or at the worst case removal of a complete array from further analysis [9–12]. After judging about the quality of the array, functional data analysis can be performed which should lead, finally to a biological conclusion.

The current paper describes a workflow for quality control and analysis of two-color microarray data. To test the proposed workflow we analyzed data obtained from an experiment setup to explore the possible mechanisms for the protective effects of dietary polyphenols on colon mucosa. A number of studies in fact demonstrated that

R. I. M. van Haaften · A. van Erk · C. T. A. Evelo (✉)
Department of Bioinformatics-BiGCaT,
Maastricht University, UNS50, Box 19,
P.O. Box 616, 6200 MD Maastricht, The Netherlands
e-mail: chris.evelo@bigcat.unimaas.nl

C. Luceri
Department of Pharmacology, University of Florence,
Florence, Italy

treatments with polyphenols had chemopreventive effects against colon carcinogenesis [13–15], probably linked to their antioxidant [16], pro-apoptotic [13] and anti-inflammatory activities.

The paper demonstrated that an insufficient quality control and not correct normalisation can lead to wrong biological conclusions.

Materials and methods

Microarray construction

The microarrays were constructed using the Rat Genome Oligo Set version 1.1 (Operon Technologies, CA, USA), composed of 70mer probes representing 5,677 well-characterized *Rattus norvegicus* genes divided into seventeen 384-wells plates. The oligonucleotides were spotted with an OmniGrid® 100 microarrayer (Genomic Solutions, Ann Arbor, MI, USA) onto poly-L-lysine glass slides (Erie Scientific Company Portsmouth, NH, USA), on the same day, using a print head with 16 pins. The Operon plates were inserted in the machine, from plate 1st to 17th, thus the oligos from every plate will end up distributed over all blocks.

Animals and samples

In the experiment, two groups of rats were compared: the control group consisted of 10 males, 5–6-week-old, Fischer 344 (F344) rats (Nossan, Correzzana, Milan, Italy) fed a high fat diet (control diet) for 2 weeks. The high fat diet was based on the AIN76 diet [17] modified to contain a high level of fat (23% corn oil w/w) and a low level of cellulose (2% w/w) to mimic the high risk of colon cancer in human populations consuming high fat diets. The experimental group consisted of 10 males, 5–6-week-old, F344 rats fed the same high fat diet as the control group, supplemented with 50 mg/kg red wine polyphenols, for 2 weeks. After killing, samples of normal colon mucosa, scraped from the connective layer with a glass slide, were harvested and placed in RNAlater (Qiagen, Milan, Italy) and stored at −80°C.

RNA isolation, labeling and hybridization

Total RNA was extracted using the RNeasy Midi kit (Qiagen, Milan, Italy). Equal amounts of RNA extracted from the colon mucosa of control diet-fed rats ($n = 10$) were pooled and used as common reference for all hybridizations.

Ten comparisons between RNAs from the 10 polyphenols-treated rats (labeled with Cy5) and the reference RNA (labelled with Cy3 (CyDye Mono-Reactive Dye Pack, Amersham, Cologno Monzese, Milan, Italy) were performed, using the indirect labeling method described by DeRisi (J. DeRisi lab, UC San Francisco, USA) (http://derisilab.ucsf.edu); for each comparison we performed an independent technical replicate (independent reverse transcriptase reaction, labeling and hybridization). The hybridization was performed at 63°C for 14–18 h.

The images were scanned using a Genepix 4000B microarray scanner (Axon Instruments, Foster City, CA, USA); the loading of the array list (to locate the reporters on the microarray) and the image analysis were performed with the GenePixPro4.1 software.

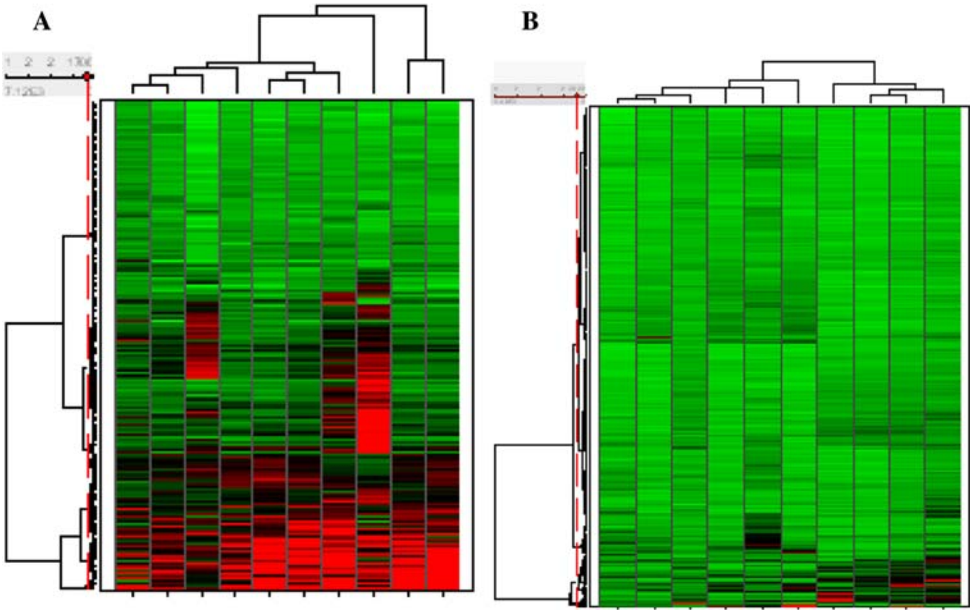

Fig. 1 **a** Hierarchical clustering after the first data analysis; genes are shown in a dendrogram based on the similarity between ten rats. **b** Hierarchical clustering after a local normalisation

On each array, “empty” spots and “not found” features were flagged automatically. Features with a strange morphology (roundness of the spot), with a clear saturated intensity status or in presence of non specific signs, like particles of dust or of dye precipitate, were flagged manually as “bad feature”.

The full dataset for this experiment was uploaded to the ArrayExpress array data repository (http://www.ebi.ac.uk/microarray-as/ae/) where it is available as experiment EMEXP-934.

Microarray standard analysis

Removal of flagged features, background subtraction and a ratio-based normalisation were performed using the Acuity 4.0 software (Axon Instruments). For each reporter, the signal log ratio (and from that, the fold change) was calculated as average of two technical replicates or as single value in presence of a missing data in the replicate.

Spotfire DecisionSite version 7.3 was used to perform a hierarchical clustering of the fold changes of the genes, of all ten animals. All genes showing a change of twofold or more in at least one experimental condition (in at least one rat) were included in the cluster analysis.

To identify biological processes affected by polyphenolic treatment, the visualization tool GenMAPP (Gene Map Annotator and Pathway Profiler, http://www.genmapp.org) version 2.0 was used. This is a generally accessible program for viewing and analyzing gene array data on microarray pathway profiles (MAPPs) representing biological pathways or any other functional grouping of genes. For GenMAPP analysis we used the Gene Ontology database (http://www.geneontology.org), the local rat MAPPS generated from the G-protein Coupled Receptor Database (http://www.gpcr.org), the KEGG database (http://www.genome.ad.jp/kegg) and MAPPs specifically designed for GenMAPP. Local MAPPs used: Rn_Contributed_20051116; gene database used: Rn-Std_20051114.gdb.

The used gene expression data were the average fold changes of the genes in the ten rats analyzed. The cut-off value for detecting a changed gene in MAPPFinder was set at 1.4 or −1.4 to point out also minor but coordinated changes.

Quality improvement and quality control of functional analyses

The raw data (intensity and background signals of both colors) and flagged features were re-plotted in a matrix that corresponds to the original array location using Spotfire DecisionSite. After re-plotting the array data in the original physical layout, it was possible to detect bad parts of the array, recognized by a non-random localization of the background signals or non-random localization of differentially expressed genes in one part of the array.

The genes changed in the pathways mostly affected by the treatment were also plotted back to the original matrix of the microarray using Spotfire DecisionSite to identify potential local-effects.

Results and discussion

After a standard analysis, cluster analysis highlighted genes (about 700) showing dissimilar patterns in 3 rats out of 10 analyzed (Fig. 1a). The differences among the expression profiles of these 3 rats cannot be assigned to the treatment, that was a short term dietary intervention with no chemical or pharmacological treatment and/or to a inter-individual variability considering that Fischer 344 are inbred rats, genetically very similar.

Functional analysis, performed analyzing these data with GenMapp/MAPPFinder, revealed the up regulation of pathways associated with cell-adhesion and oxidative stress (see Table 1). These results were in contrast with biological data: previous studies performed in our lab in fact demonstrated a strong antioxidant effect of polyphenolic treatments on rat colon mucosa [16, 18].

Visualization of the signal log ratios in a matrix that corresponds to the original array location we observed that in three hybridizations the ratios were not randomly spread

Table 1 Results of the GenMAPP/MappFinder analysis of pathways affected by red wine polyphenols

	Number of genes changed
Block by block normalisation	
Pathways down-regulated	
Rn_Prostaglandin_synthesis_regulation	13
Rn_MAPK_signaling_pathway_ KEGG	46
Rn_Oxidative_Stress	9
Rn_TGF_Beta_Signaling_Pathway	20
Rn_Cytokines_and_Inflammatory_Response_Biocarta	15
Pathways up-regulated	
Rn_G1_to_S_cell_cycle_Reactome	3
Rn_Cell_cycle_KEGG	4
Global normalisation	
Pathways down-regulated	
Rn_MAPK_signaling_pathway_ KEGG	21
Rn_TGF-beta-Receptor_NetPath_7	15
Pathways up-regulated	
Rn_Focal_adhesion_KEGG	5
Rn_EGFR1_NetPath_4	4
Rn_Oxidative_Stress	5

All pathways with a $P < 0.05$ are shown for pathway enrichment across the array: one block, out of the 16 printed on the array, contains in fact genes with a high signal log ratio. Moreover, re-plotting genes belonging to the cell–cell adhesion and oxidative stress pathways, back to the original matrix of the microarray, we observed that they were mainly located in the same block (Fig. 2).

The disagreement between microarray and biological data was therefore due to a non-random distribution of the signal log ratios across the array. Such effects could be caused by irregularities in the spotting procedure leading to high background values compared to intensity signals or by a high print tip variability (each block is printed by a different pin). In the present case, in the arrays used to analyze the RNA of three rats presented a block with a very low signal (but not low enough to be called "not found"). It is interesting to note that despite the randomized spotting of the oligos into the array, there is a chance that a large fraction of genes involved in the same pathways end up in the same block.

In the current example it was enough to replace the standard global ratio-based normalisation with a lowest block-by-block normalisation to remove the origin of the artifact.

After the quality improvement and the block-by-block (lowest) normalisation, the hierarchical clustering showed that all ten rats looked similar to each other (Fig. 1b); functional analysis identified as biological processes down-regulated by the red wine polyphenols, the oxidative stress, together with other pathways, not identified by the previous analysis such as the prostaglandin synthesis regulation and the cytokines and inflammatory response (Table 1) .

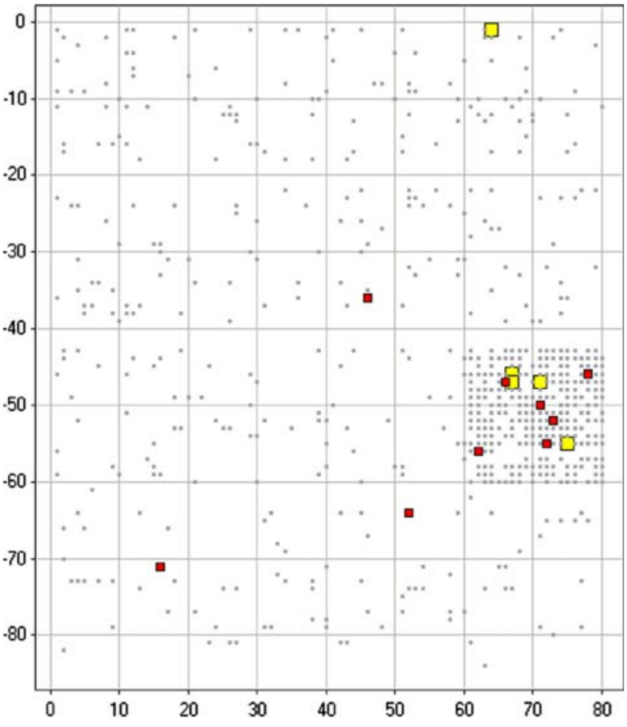

Fig. 2 Localization on the microarray of genes differentially expressed involved in cell adhesion (*yellow squares*) and oxidative stress (*red circles*). Almost all these genes are located in the same block. The *axes* represent the rows and the columns of the microarray

General workflow for quality control and quality improvement

The approach used to analyze the biological experiment described in this manuscript can be summarized in a general workflow for quality control and quality improvement. The workflow consists of different steps starting with the removal of flagged features, the background subtraction and a global normalisation.

The second step is a hierarchical clusterization to visualize the expression profiles of the experimental groups. The cluster analysis can suggest the presence of biological differences among groups/rats. If these differences are not supported or even in contrast with biological results, we suggest, as third step, the visualization of microarray data to identify the presence of technical artifacts: at this point there are three possibilities (1) the quality of the complete array is bad; no further analysis is possible; (2) the quality of part of the array is bad; quality improvement of the array is possible; (3) the quality of the array is acceptable for further functional analysis. In the second case a possible step in the workflow is a new local normalisation. When the quality of the array is finally satisfactory, the next step in the workflow is a functional analysis. After that a re-plot of the genes involved in pathways found to be modulated, back to the original matrix of the microarray, can reveal or exclude any "local effects".

The suggested workflow allows the improvement of microarray analysis, through an integration of extensive physical data observation, pathway analysis, clustering and dedicated normalisation procedures.

Acknowledgments We thank the European Nutrigenomics Organisation (NuGO) who provided an exchange grant for Dr. Luceri. We thank Drs. Simona Toti and Cinzia Castagnini for the critical review of the manuscript.

References

1. Jourdren L, Le Crom S (2005) Doelan: a solution for quality control monitoring of microarray production. Bioinformatics 21:4194–4195

2. Hessner MJ, Meyer L, Tackes J, Muheisen S, Wang X (2004) Immobilized probe and glass surface chemistry as variables in microarray fabrication. BMC Genomics 5:53
3. Holloway AJ, van Laar RK, Tothill RW, Bowtell DD (2002) Options available—from start to finish—for obtaining data from DNA microarrays II. Nat Genet 32(Suppl):481–489
4. Bylesjö M, Sjödin A, Eriksson D, Antti H, Moritz T, Jansson S, Trygg J (2006) MASQOT-GUI: spot quality assessment for the two-channel microarray platform. Bioinformatics 22:2554–2555
5. Wang X, Ghosh S, Guo SW (2001) Quantitative quality control in microarray image processing and data acquisition. Nucleic Acids Res 29:E75
6. Burgoon LD, Eckel-Passow JE, Gennings C, Boverhof DR, Burt JW, Fong CJ, Zacharewski TR (2005) Protocols for the assurance of microarray data quality and process control. Nucleic Acids Res 33:e172
7. Carter DE, Robinson JF, Allister EM, Huff MW, Hegele RA (2005) Quality assessment of microarray experiments. Clin Biochem 38:639–642
8. Dumur CI, Nasim S, Best AM, Archer KJ, Ladd AC, Mas VR, Wilkinson DS, Garrett CT, Ferreira-Gonzalez A (2004) Evaluation of quality-control criteria for microarray gene expression analysis. Clin Chem 50:1994–2002
9. Sauer U, Preininger C, Hany-Schmatzberger R (2005) Quick and simple: quality control of microarray data. Bioinformatics 21: 1572–1578
10. Buness A, Huber W, Steiner K, Sültmann H, Poustka A (2005) ArrayMagic: two-colour cDNA microarray quality control and preprocessing. Bioinformatics 21:554–556
11. Tran PH, Peiffer DA, Shin Y, Meek LM, Brody JP, Cho KW (2002) Microarray optimizations: increasing spot accuracy and automated identification of true microarray signals. Nucleic Acids Res 30:e54
12. Wang X, Hessner MJ, Wu Y, Pati N, Ghosh S (2003) Quantitative quality control in microarray experiments and the application in data filtering, normalization and false positive rate prediction. Bioinformatics 19:1341–1347
13. Caderni G, De Filippo C, Luceri C, Salvadori M, Giannini A, Biggeri A, Remy S, Cheynier V, Dolara P (2000) Effects of black tea, green tea and wine extracts on intestinal carcinogenesis induced by azoxymethane in F344 rats. Carcinogenesis 21:1965–1969
14. Luceri C, Caderni G, Sanna A, Dolara P (2002) Red wine and black tea polyphenols modulate the expression of cycloxygenase-2, inducible nitric oxide synthase and glutathione-related enzymes in azoxymethane-induced F344 rat colon tumors. J Nutr 132:1376–1379
15. Femia AP, Caderni G, Vignali F, Salvadori M, Giannini A, Biggeri A, Gee J, Przybylska K, Cheynier V, Dolara P (2005) Effect of polyphenolic extracts from red wine and 4-OH-coumaric acid on 1, 2-dimethylhydrazine-induced colon carcinogenesis in rats. Eur J Nutr 44:79–84
16. Giovannelli L, Testa G, De Filippo C, Cheynier V, Clifford MN, Dolara P (2000) Effect of complex polyphenols and tannins from red wine on DNA oxidative damage of rat colon mucosa in vivo. Eur J Nutr 39:207–212
17. Ad hoc Committee on Standards of Nutritional Studies (1997) Report of the American Institute of Nutrition. J Nutr 107:1340–1348
18. Dolara P, Luceri C, De Filippo C, Femia AP, Giovannelli L, Caderni G, Cecchini C, Silvi S, Orpianesi C, Cresci A (2005) Red wine polyphenols influence carcinogenesis, intestinal microflora, oxidative damage and gene expression profiles of colonic mucosa in F344 rats. Mutat Res 591:237–246