RESEARCH ARTICLE

Quality Control Usage in High-Density Microarrays Reveals Differential Gene Expression Profiles in Ovarian Cancer

Vanessa Villegas-Ruiz[1,2], Jose Moreno[3], Karina Jacome-Lopez[1], Alejandro Zentella-Dehesa[4,5], Sergio Juarez-Mendez[1]*

Abstract

There are several existing reports of microarray chip use for assessment of altered gene expression in different diseases. In fact, there have been over 1.5 million assays of this kind performed over the last twenty years, which have influenced clinical and translational research studies. The most commonly used DNA microarray platforms are Affymetrix GeneChip and Quality Control Software along with their GeneChip Probe Arrays. These chips are created using several quality controls to confirm the success of each assay, but their actual impact on gene expression profiles had not been previously analyzed until the appearance of several bioinformatics tools for this purpose. We here performed a data mining analysis, in this case specifically focused on ovarian cancer, as well as healthy ovarian tissue and ovarian cell lines, in order to confirm quality control results and associated variation in gene expression profiles. The microarray data used in our research were downloaded from ArrayExpress and Gene Expression Omnibus (GEO) and analyzed with Expression Console Software using RMA, MAS5 and Plier algorithms. The gene expression profiles were obtained using Partek Genomics Suite v6.6 and data were visualized using principal component analysis, heat map, and Venn diagrams. Microarray quality control analysis showed that roughly 40% of the microarray files were false negative, demonstrating over- and under-estimation of expressed genes. Additionally, we confirmed the results performing second analysis using independent samples. About 70% of the significant expressed genes were correlated in both analyses. These results demonstrate the importance of appropriate microarray processing to obtain a reliable gene expression profile.

Keywords: High density microarrays - gene expression profiles - quality control - ovarian cancer.

Asian Pac J Cancer Prev, **17 (5)**, 2519-2525

Introduction

Ovarian cancer is the sixth most common malignancy in México represent the fourth leading cause of gynaecological cancer death (Jemal et al., 2008), despite advanced in clinical diagnosis, the survival to five years is less to 50 %. Microarray gene expression assays (MGEA) are a powerful tool for the simultaneous analysis of thousands of genes. Essentially, the number of studies reported in the ARRAYEXPRESS web site has reached 63517 experiments and 1,912,232 assays. In a short time span, genome studies have been applied to great diversity of biological models, including human diseases. Moreover, these assays have allowed the definition of gene expression profiles-associated (Ma et al., 2009; Rong et al., 2013).

MGEA and next generation sequencing has been applied for survey molecular signatures in several types of cancer, including: lymphoma (Knudsen et al., 2015), colon (Dinalankara and Bravo, 2015; Sun et al., 2015), cervical (Horikawa et al., 2015), head and neck (Placa et al., 2015), breast (Li et al., 2015), ovarian (Cai et al., 2015), among others, providing potential molecular markers for clinical application in diagnosis and prognosis. Additionally, cancer gene signature which is used on system biology approach, could allow visualization and understanding of the disease complexity, based on the prediction of interactions among protein-protein, DNA-protein, RNA-DNA, RNA-protein (Panteris et al., 2007; Sonachalam et al., 2012).

Gene expression microarrays have presented a great evolution becoming a solid and robust platform for study gene interactions. A quality assurance plan (QAP) describes steps for microarray processing (array production, hybridization, and data analysis) and is focused in the integration of information to identify low

*[1]Experimental Oncology Laboratory, Research Department, National Institute of Pediatrics, [2]Genetics Department-Research Unit, Institute of Ophthalmology, "Conde de Valenciana", [3]Research Direction, Juarez Hospital of Mexico, [4]Medicine Genomic and Environmental Toxicology Department, Biomedical Research Institute, UNAM, [5]Unit of Biochemistry, National Institute of Medical Sciences and Nutrition "Salvador Zubirán", Mexico D.F., Mexico *For correspondence: ser.mend@gmail.com*

quality in production (Burgoon et al., 2005). (http://www.stat.duke.edu/~mw/ABS04/RefInfo/expression_ever_manual.pdf). In laboratory assays, several factors could affect MGEA by systematic error including: sample preparation, hybridization and scanning. Additionally, the variability of each procedure could affect the integrity data for meaningful interpretation that support conclusion on futures analysis, hypothesis generation, among others (Burgoon et al., 2005).

Several efforts have been made for identified low quality microarrays, including technical, biological replicate and visualization in terms of sample clustering (Zakharkin et al., 2005). Algorithms and specialized genomics software have been developed for normalization and background correction by data comparison (Quackenbush, 2001).

Quality controls in microarrays provide indirect inspection of microarray processing, McCall et al, identifies that approximately 10% of microarray public data are poor in quality (McCall et al., 2011). Until this moment it is not clear the impact in gene expression profiles based on poor quality controls microarrays.

In this study we used the Affymetrix U133_Plus_2 array with two controls (Poly A and hybridization) for processing in order to allow the identification of correct microarray labelling; however, the experimental variation and impact in over and under estimation gene signature is unknown. We examine with in silico approach the impact of microarray quality control in gene expression signature using public microarrays data of ovarian healthy tissue, ovarian cancer and ovarian cell lines. We used microarrays quality controls as screening data available in public repositories, reducing underestimation/overestimation of gene expression profiles by means of poor microarrays quality controls (MQC). MCQ Analysis showed ~40 of available data are poor in Quality Controls (QC). Additionally, the uses of microarray that have any errors in quality control affect gene expression profile.

Materials and Methods

Database Selection

We performed a meta analysis of microarray data, we selected in public repository all available data of *Homo sapiens*, ovarian gene expression tissue by means of U133 Plus_2 Affymetrix microarray (EMBL-EBI, GEO) 47,000 transcripts and mRNA variants of which ~29,000 were correctly annotated in this array.

The arrays are comprised of eukaryotic hybridization controls bioB, bioC, bioD from biotin synthesis of E. coli and cre from P1 bacteriofago, the final concentration is at 1.5 pM, 5 pM, 25 pM and 100pM, respectively; eukaryotic Poli-A RNA control (labelling controls; dap, thr, phe and lys) from B. subtilis at final concentration in the sample: 1:7,500, 1:25,000, 1:50,000 and 1:100,000, respectively; 100 probes set to normalization and housekeeping genes (GADPH, beta-Actin, ISGF-3 (STAT1), among others). The microarray data were downloaded form the webpage Gene Expression Omnibus (GEO) (Edgar et al., 2002) http://www.ncbi.nlm.nih.gov/geo/, and ARRAYEXPRESS data base (Brazma et al., 2003; Rocca-Serra et al., 2003; Parkinson et al., 2005) http://www.ebi.ac.uk/arrayexpress/. We included in this study data from microarray performed from 2007 to 2012 based on 31 publications. Although there were a total of 2300 arrays achieved, we included 540 assays in this study.

We downloaded the U133_Plus_2 Affymetrix microarray assays of epithelial ovarian cancer in advances stages (lll and lV), ovarian cancer cell lines and healthy ovarian tissue. All the samples selected were untreated. From tumours and cell lines that fulfilled the minimum parameters, we excluded microarray data as if it did not have the previous characteristics.

Of these, we were only able to examine 464 as the remaining 76 presented downloading errors that kept us from having an appropriate data collection. For analysis, microarray assays were sub-classified as follows: ovarian cell lines (OCL=72), epithelial serous ovarian tumours (ESOT=308), and healthy ovarian tissue (HOT=84).

Quality control analysis

Microarray Quality Control (MQC) Analysis (MQCA) was performed using Affymetrix Expression console. This software is available in Affymetrix website http://www.affymetrix.com/estore/browse/level_seven_software_products_only.jsp?productId=131414#1_1. The validity of all chosen experiments was independently assessed by means of the MAS5 Statically Algorithm, Probe logarithmic Intensity Error Estimation (PLIER), and Robust Multichip Average (RMA). Analysis was performed by means of Affymetrix Expression Console software v1.4.1 for microarray quality controls. We used a standard set of defaults has been provided.

Quality controls data was examined for outliers when compared to other highly related samples by means of Hybridization (Spike Controls), labelling (Poly-A RNA controls), internal control genes (housekeeping), and global array. The line metrics were based on hybridization and the labelling controls intensity signal. The histogram signal was performed using 100 quality controls probe set and Pearson´s correlation determined the relationship between arrays (r2).

Gene expression analysis

The differential gene expression analysis were archived by means of the Partek Genomics Suite v6.5 (Partek Incorporated, saint Luis, MO). We used the Robust Multichip Average (RMA) for background correction, Quantile Normalization, Median Polish Sumarization, and data were Log2 transformed, according a previous studies (Juarez et al., 2013). The data were categorized and grouped by means of Principal Component Analysis (PCA). The differential expressed genes were detected using Analysis of Variance (ANOVA) with HOT as reference control (base line). Moreover, OCL and ESOT were compared against a reference tissue using geometric least squares means model. The significant and differential expressed genes were selected by means of cutoff; fold change (>3 and <-3) and False Discovery Ratio (FDR) <0.005. Finally, the relationship among genes that expressed significantly different was visualized by means of a Venn diagram and clustering.

Results

Data selection

We downloaded a total of 31 experiments that contain several microarrays assays. Nevertheless, only 540 assays used and categorized as follows: healthy ovarian tissue, ovarian cancer cell lines and untreated epithelial serous ovarian tumors stages III and IV. In microarray examination only 464 assays were evaluated, the remaining 76 showed downloading errors that could not allow us to have an appropriate data collection, the data collection are shown in Additional File 1. For reproducibility the study was fractionated. We performed the first analysis with available data performed form 2007 to march 2010. From the latter selection, 283 microarray assays were including in the analysis. For analysis, the assays were classified as follows: HOT = 46, ESOT = 201 and OCL = 36.

Microarray quality controls analysis (MQCA) is essential for inter-groups comparisons

The MQCA of all assays are depicted in Figure 1A-1B, indicating the presence of major differences in the proportions of internal labelling controls unlike the hybridization controls Figure 1C. Moreover, the MQC of each microarray showed distribution differences and correlations. These results suggest some errors during microarray processing affect the global array expression. Consequently, the validity of each experiment was independently assessed by means of RMA, MAS5 and Plier analysis algorithms using Affymetrix Expression Console software v1.4.1. The coincidence in MQCA is depicted in Venn diagrams; the plots show the correlation of the microarray noises. The data was analysed independently by each algorithm Figure 2.

Another two Plier and MAS5 also identified all 55 RMA failures; still Plier picked 37 more (total=92) and MAS5 picked 56 more (total=111). Finally, the level of coincidence between Mas5 and Plier was 70% (84 of 119). Although these results suggest that MAS5 is the most sensitive algorithm tool for selecting useful microarray results, the best accuracy appears to be achieved by RMA. This suggests that it is important to examine which of the algorithms reproduces the data rejection obtained by MQCA parameters. Moreover, the arrays were individually rejected by each of the algorithms as well as by different combinations as follows: RMA only = 0, Plier only = 8, Mas5 only = 27, RMA+Plier = 0, RMA+Mas5 = 0, Plier+Mas5 = 29, RMA+Plier+Mas5 = 55 Figure 2.

Up to here, the data indicate that the MQCA with expression console software allow the identification of microarray files with successful MQC (MQCS, N=164), and non-successful MQC (MQCNS, N= 119). Additionally, MQCS and MQCNS files were examined independently and categorized based on tissue type (HOT=13, ESOT=131, OCL=20 for QCMS; and OH=33, ESOT=70, OCL=16 for QCMNS). As we mentioned above, the analysis was performed only with MQCS files. Confirming the result obtained with the unfractionated MQCS samples; we did not observe

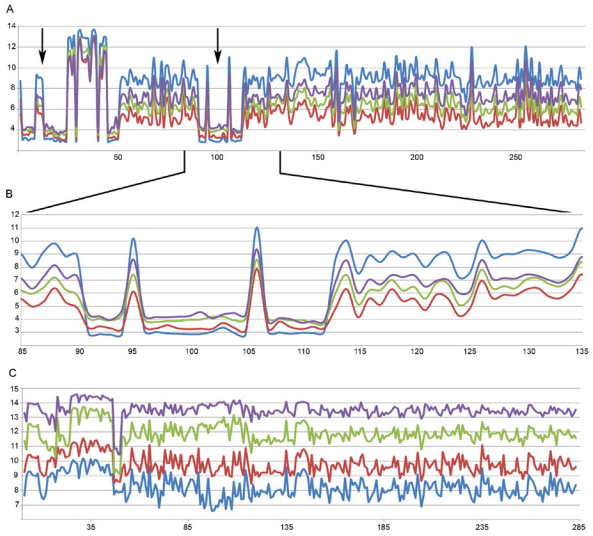

Figure 1. Microarray Quality Control Analysis

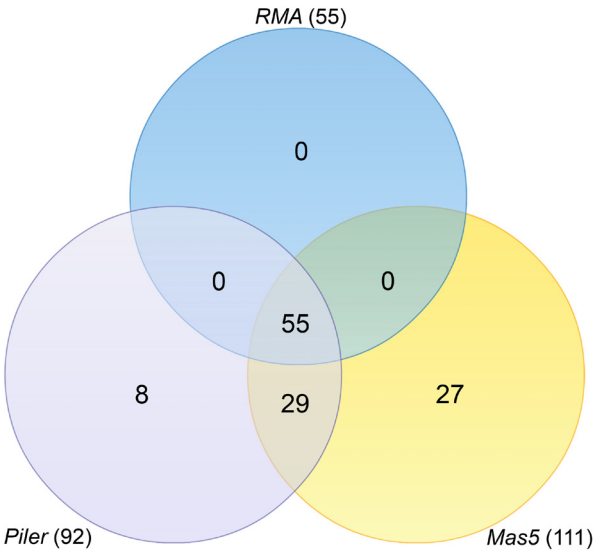

Figure 2.

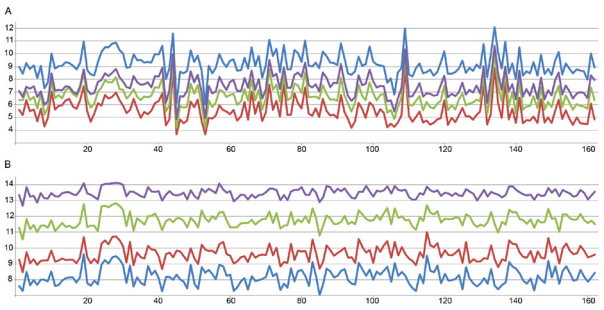

Figure 3. ??

changes in hybridization and labelling controls Figure 3. We also observed low dispersion among microarray files as follows on histogram. As well, Pearson's correlation shows more equality in the data. On the first review, we could show several changes on the MQC proportion in some microarrays files; the data presented heterogeneity and low microarray correlation, these data suggest low technical reproducibility.

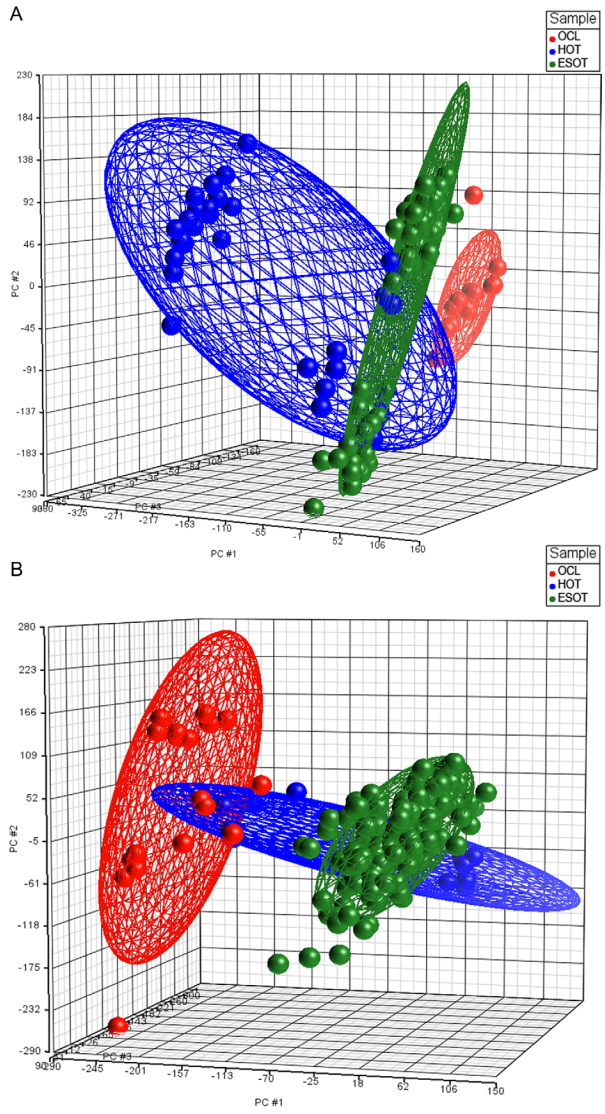

Figure 4. ??

Microarray Quality Control affected the sources of gene expression profiles

We noticed that QCA showed several errors in public microarray data, we wanted to know if there were gene expression profiles MQCA-associated. In this inspection we used the same parameters on cut-off and statics but we employed the Partek Genomic Suite v6.6 software for microarrays analysis.

The first step was to categorize the data by means of MCQA (MCQS and MCQNS) after we had visualized the distribution of quality controls including the ones in microarray. We noticed that clearly the distributions of markers in MQCS samples were undifferentiated, unlike QCMNS as we expected according to previous analysis of MCQA.

In order to explore gene expression MQCA-based, the collection microarray data were classified in three groups as follows: MQCS (undifferentiated microarray data in quality controls analysis), MQCNS (microarray data that showed inappropriate quality control analysis) and AF (including all microarray data in the study). The data was analysed independently and sub-classified tissue-based (HOT, ESOT and OCL). The Principal Component Analysis (PCA) revealed unspecific grouping for each condition in MQCNS files Figure 4a unlike to MQCS files; each tissue lineage was clustering tissue-associated Figure 4b.

Noises microarray data generates differential gene expression profile

MQCA and PCA have revealed differences between MCQNS and MCQS microarray files, showing several changes in QC plot, high dispersion and low correlation Figure 1A-B, 4a, Additional File 3A, respectively. The next step was to identify global gene expression on AF, MQCNS and MQCS. In order to identified gene expression profiles based on quality control analysis. The differential expressed genes were identified by contrast between ESOT and OCL versus HOT as base line expression. Significant genes were selected as follows in the methods section. The analysis based in QC showed heterogeneous data on the number of significant expressed genes profile for three analysis with major differences in gene expression changes for MQCNS files (N= 2412), followed by MQCS (N=1815) and AF (N=1788).

The gene expression profile agreement in three analyses was as follows: QCMNS (N =2412), MQCS (N = 1815), AF (N =1788), MQCNS only (N = 878), MQCS only (N = 808), AF only (N = 42). The correlation gene expression profile revealed mainly differences among MQCNS, MQCS and AF was a follows: MQCNS + MQCS (N=12), MQCNS + AF (N=751), MQCS + AF (N= 224), MQCNS + MQCS + AF (N=771).

Reproducible of gene expression profiles using MQCA

We performed independent analysis with the second selection of microarray files. These data were performed from April 2010 to November 2012. A total 15 experiments were identified; in this selection 181 microarrays files were used. The data was sub classified as follows: OCL (N=36), ESOT (N= 107), and HOT (N = 38).

Like in our previous analysis these data were analysed by expression console software. Surprisingly in this analysis we could identify more coincidence in inappropriate data analysed as follows: RMA (N=107), MAS5 (N=117) and Plier (N=119). The correlation using the algorithms was: RMA only =0, Plier only =8, MAS5 only = 5, RMA+Plier = 0, RMA+MAS5 =0, Plier+MAS5 =1, RMA+Plier+MAS5 = 101. The second analysis showed that all HOT and OCL files have been wrong in MQCA, only 66 EOT files did not present errors using the RMA, MAS5 and Plier algorithms.

In this analysis only ~36% of available data were consider successful in MQCA and all data belonged to the group of ESOT. In order to identify correlation in differential gene expression of our first and second analysis, we compared HOT versus ESOT of the first and second analysis performed. The HOT samples were used as a base line (N=13) for both analyses. The First Analysis (FA) was included as follows: (ESOT, N=131 vs. HOT, N=13) and in Second Analysis (SA) (ESOT, N= 66 vs OH, N=13). Significant and differential expressed genes

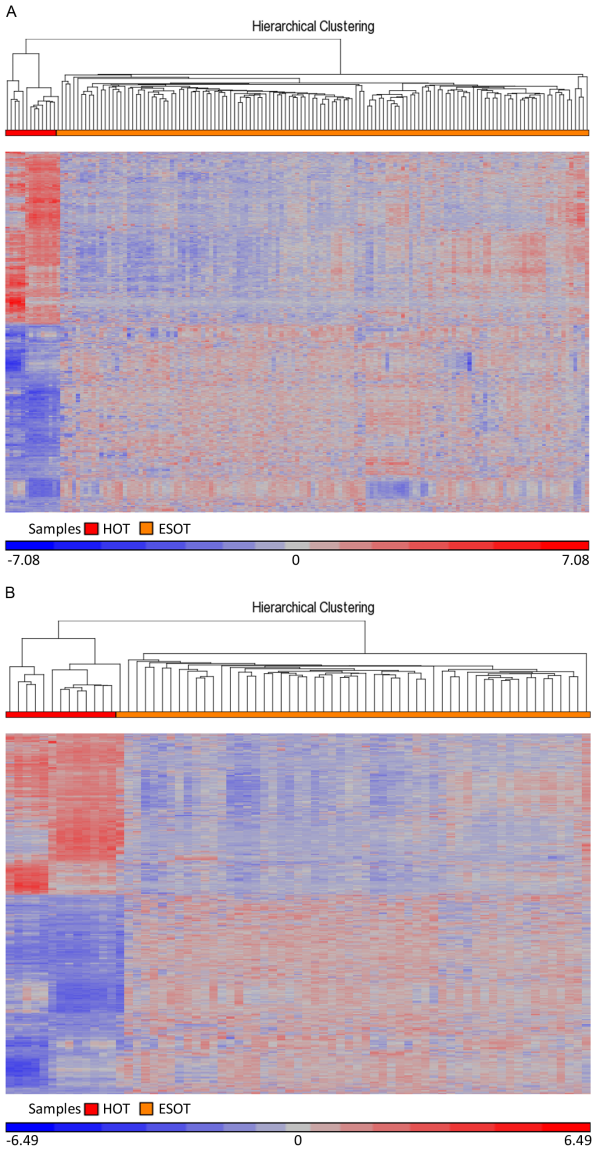

Figure 5. ??

were selected of previous parameter as already described in the methods.

The analysis showed 1613 and 1947 differential expressed genes in FA and SA respectively. These results showed a correlation of ~68 % that corresponds to 1445 significantly expressed genes. Additionally, the significant expressed genes were clustering Figure 5.

Discussion

Cancer is a complex disease, it is constituted by several characteristics including: the evading of apoptosis, self-sufficiency in growth signals, limitless replicative, potential tissue invasion and metastasis, among others (Hanahan and Weinberg, 2000; Hanahan and Weinberg, 2011). The hallmarks of cancer are based on high complexity of molecular interaction including DNA-DNA, DNA-RNA, DNA-Proteins, RNA-RNA, RNA-Proteins, and Proteins-Proteins. One of the greatest challenges in biomedical sciences is the identification of molecular markers such as: prognosis, diagnosis and treatment. Several strategies have been employed in ovarian cancer research; one of them is high throughput technology. The gene expression profiles in cancer allow the selection of different tissue/tumour types as well as the potential prediction of drug responses, diagnosis and prognosis (Chang et al., 2011; Heyn and Esteller, 2012).

The high throughput technology allowed analyses of transcriptomes in a single experiment as well as the detection of gene expression profiles as a landscape of molecular biology in cancer, this technology has been employed to elucidate malignance-associated deregulated genes (Chang et al., 2011). The microarray studies have been increasing considerably allowing the constitution of repositories databases as ARRAYEXPRESS and GEO (Edgar et al., 2002; Brazma et al., 2003; Rocca-Serra et al., 2003; Parkinson et al., 2005; Kolesnikov et al., 2015); in ARRAYEXPRESS for example there are at the moment 39,074 experiments and 1,121,411 assays. It is well known in microarray studies that there are many sources that affect the values of gene expression. The quality controls are the references of success hybridization and microarray processing. Inconspicuous erroneous in quality control could affect gene signal as evidenced at the results of McCall et al., which showed that approximately 10% of public available array are poor in quality controls. Strikingly, our results showed 40% approximately of public data (Affymetrix U133_Plus_2) are poor in quality controls, based on RMA, MAS5 and Plier algorithms Figure 1A-B and 2.

The quality controls are essential to examine microarray processing;, housekeeping, Eukarotic Poli-A RNA control (B. subtilis: lys, phe, thr, dap) and hybridization controls (bioB, bioC, bioD and cre) constitute Affymetrix U133_Plus_2 microarray. The probe set is based on oligo designs of 25 nucleotides, constituted by perfect-match (PM) and mismatch probes (MM). RMA is the most popular algorithm used on microarray analysis, although, this algorithm only examine PM probes set (Irizarry et al., 2003). The MQCA based on RMA showed to have the most accurate algorithms; the data obtained in this study was consistent whit Plier and Mas5 analyses. Nevertheless, the best sensitive algorithms were MAS5 as its analysis showed more noise files than RMA and Plier algorithms (~70%). This results from MAS5 and Plier can combine the signal from multiples Perfect-Match and Mismatch values (Seo et al., 2006; Pepper et al., 2007).

Microarray processing is based on several steps such as: RNA amplification (in vitro transcription), reverse transcription, labelling, fragmentation and hybridization. Labelling and hybridization controls evaluate the microarray processing. Our results showed several changes in labelling controls Figure 1A-B, suggesting inappropriate microarray processing with a consequently low reliability in the results.

The most critical deterrent to success in microarray assay is high level of technical noises; the MQCA are set to identify noises in assays, but microarrays with no success as could be QC could have relative false measurements in gene expression. The human transcriptome is examined by microarray assays revealing gene expression profile, however if the processing is incorrect the results may

be altered reducing the correlation among experiments (Additional File 3A) and sample distribution (Additional File 2A and Additional File 4B). PCA was performed based on analysis of MQCA, revealing files with success as well as with no success in microarray possessing, the dimensional space were more specific for microarray successful in MQCA Figure 4B unlike microarray non-successful Figure 4A. These results suggest that each group has different gene expression profile because level of expression in the sample was not reduced into a small set correlated PCA (Bicaku et al., 2012). It is important to include these parameters for the variability in the data analysed.

Affymetrix microarrays are high density as they use 25 mer oligonucleotide in 8 μm2 by probe, the probe set design provided high sensitivity and specificity hybridization and the probe set density cover all transcriptome. However, as its scale could have many factors affecting the microarray assays, it can present variations in the relative values of gene expression (Seo et al., 2006). We think that QC is the most sensitive measurement of experimental variation in microarrays processing. Our results showed low correlated in significantly and differentially expressed genes based on MQCA (771) and high number of genes were exclusive of noises files (878) and success files (808) data no show. These results suggest that several over and under estimates genes could be detected. The reduction of noises samples could generate solid groups of comparison (Additional File 3B).

On the other hand, we could observe high reproducibility (68 %) using microarrays that that not have inappropriate quality controls, although it was not possible to identify files of OCL and HOT in second analysis, the gene expression profile found in tumors were clustering clustering Figure 5. It is possible that some variations in gene expression profiles in FS (168 genes) and SA (502 genes) were observed by tumors heterogeneity, probably.

In conclusion, there are many algorithms for microarray correction, the use of the noises files processed affect the results underestimating or overestimate genes that would impact our understanding of biological process. Our results showed that approximately 40% of microarray data available in public repositories, is poor in QC. We recommend the use of microarray file that no have alteration in quality controls.

Acknowledgements

This work was supported in part by grant Basic Science CONACyT 243233 and Federal Founds, National Institute of Pediatrics. We finally thank PhD. Hector Pedro Salazar Garcia, MC. Silvina Romero and MD. Erika Paulina Lopez for critical review the manuscript.

References

Bicaku E, Xiong Y, Marchion DC, et al (2012). In vitro analysis of ovarian cancer response to cisplatin, carboplatin, and paclitaxel identifies common pathways that are also associated with overall patient survival. *Br J Cancer*, **106**, 1967-75.

Brazma A, Parkinson H, Sarkans U, et al (2003). ArrayExpress--a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res*, **31**, 68-71.

Burgoon LD, Eckel-Passow JE, Gennings C, et al (2005). Protocols for the assurance of microarray data quality and process control. *Nucleic Acids Res*, **33**, e172.

Cai SY, Yang T, Chen Y, et al (2015). Gene expression profiling of ovarian carcinomas and prognostic analysis of outcome. *J Ovarian Res*, **8**, 50.

Chang C, Wang J, Zhao C, et al (2011). Maximizing biomarker discovery by minimizing gene signatures. *BMC Genomics*, **12 Suppl 5**, S6.

Dinalankara W, Bravo HC (2015). Gene expression signatures based on variability can robustly predict tumor progression and prognosis. *Cancer Inform*, **14**, 71-81.

Edgar R, Domrachev M, Lash AE (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*, **30**, 207-10.

Hanahan D, Weinberg RA (2000). The hallmarks of cancer. *Cell*, **100**, 57-70.

Hanahan D, Weinberg RA (2011). Hallmarks of cancer: the next generation. *Cell*, **144**, 646-74.

Heyn H, Esteller M (2012). DNA methylation profiling in the clinic: applications and challenges. *Nat Rev Genet*, **13**, 679-92.

Horikawa N, Baba T, Matsumura N, et al (2015). Genomic profile predicts the efficacy of neoadjuvant chemotherapy for cervical cancer patients. *BMC Cancer*, **15**, 739.

Irizarry RA, Bolstad BM, Collin F, et al (2003). Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res*, **31**, e15.

Jemal A, Siegel R, Ward E, et al (2008). Cancer statistics, 2008. *CA Cancer J Clin*, **58**, 71-96.

Juarez-Mendez S, Zentella-Dehesa A, Villegas-Ruiz V, et al (2013). Splice variants of zinc finger protein 695 mRNA associated to ovarian cancer. *J Ovarian Res*, **6**, 61.

Knudsen S, Hother C, Gronbaek K, et al (2015). Development and blind clinical validation of a microRNA based predictor of response to treatment with R-CHO(E)P in DLBCL. *PLoS One*, **10**, e0115538.

Kolesnikov N, Hastings E, Keays M, et al (2015). ArrayExpress update--simplifying data submissions. *Nucleic Acids Res*, **43**, D1113-6.

Li MH, Fu SB, Xiao HS (2015). Genome-wide analysis of microRNA and mRNA expression signatures in cancer. *Acta Pharmacol Sin*, **36**, 1200-11.

Ma XJ, Dahiya S, Richardson E, et al (2009). Gene expression profiling of the tumor microenvironment during breast cancer progression. *Breast Cancer Res*, **11**, R7.

McCall MN, Murakami PN, Lukk M, et al (2011). Assessing affymetrix GeneChip microarray quality. *BMC Bioinformatics*, **12**, 137.

Panteris E, Swift S, Payne A, et al (2007). Mining pathway signatures from microarray data and relevant biological knowledge. *J Biomed Inform*, **40**, 698-706.

Parkinson H, Sarkans U, Shojatalab M, et al (2005). ArrayExpress--a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res*, **33**, D553-5.

Pepper SD, Saunders EK, Edwards LE, et al (2007). The utility of MAS5 expression summary and detection call algorithms. *BMC Bioinformatics*, **8**, 273.

Placa JR, Bueno Rde B, Pinheiro DG, et al (2015). Gene expression analysis of laryngeal squamous cell carcinoma. *Genom Data*, **5**, 9-12.

Quackenbush J (2001). Computational analysis of microarray data. *Nat Rev Genet*, **2**, 418-27.

Rocca-Serra P, Brazma A, Parkinson H, et al (2003).

ArrayExpress: a public database of gene expression data at EBI. *C R Biol*, **326**, 1075-8.

Rong G, Kang H, Wang Y, et al (2013). Candidate markers that associate with chemotherapy resistance in breast cancer through the study on Taxotere-induced damage to tumor microenvironment and gene expression profiling of carcinoma-associated fibroblasts (CAFs). *PLoS One*, **8**, e70960.

Seo J, Gordish-Dressman H, Hoffman EP (2006). An interactive power analysis tool for microarray hypothesis testing and generation. *Bioinformatics*, **22**, 808-14.

Sonachalam M, Shen J, Huang H, et al (2012). Systems biology approach to identify gene network signatures for colorectal cancer. *Front Genet*, **3**, 80.

Sun QL, Zhao CP, Wang TY, et al (2015). Expression profile analysis of long non-coding RNA associated with vincristine resistance in colon cancer cells by next-generation sequencing. *Gene*, **572**, 79-86.

Zakharkin SO, Kim K, Mehta T, et al (2005). Sources of variation in Affymetrix microarray experiments. *BMC Bioinformatics*, **6**, 214.