Author's response to reviews

Title: Gene Expression Profile of Normal Human Mammary Epithelial Cells Following Exposure to Oxythioquinox as Monitored by DNA Microarrays

Authors:

Maureen R Gwinn (mwg9@cdc.gov)
Diana L Whipkey (drw9@cdc.gov)
Ainsley Weston (agw8@cdc.gov)

Version: 2 Date: 27 July 2004

Author's response to reviews: see over
Kristian Almstrup:
The paper by Gwinn et al. entitled “Gene Expression Profile of Normal Human Mammary Epithelial Cells Following Exposure to Oxythioquinox as Monitored By DNA Microarrays” describes changes in gene expression in primary NHMECs cells after exposure to Oxythioquinox OTQ. The authors have looked both at p53 protein levels, cell viability, after different time points of OTQ exposure in four NHMECs cell strains. They then have tried to identify gene expression markers of the exposure using Affymetrix GeneChips and tried to verify some results with real time PCR. The study is well written but with some statements repeated several times.

Major concerns:
The study is performed with good controls but I miss to see the described preliminary results about the choice of 6.25 microM OTQ and 2 hours exposure as references mentioned (2 and 3) are rather old. What I really miss to see is some dose-response relations of OTQ exposure.

**Dose-response data was performed mainly with immunohistochemistry for p53 proteins levels and trypan blue analysis for cell viability. The ideal concentration for this study was one that showed an increase in p53 levels with no decrease in cell viability. The majority of references on OTQ are fairly old—most studies were performed when it was first in use in the seventies. We have made a note of this and added it to the Methods section, p 6.**

SOM clustering rely on some prior information about the data maybe from a principal component analysis (PCA) in order to choose the most suitable geometric configuration for the clusters and the number of partitions into which the data is to be divided. It remains unclear what criteria the authors have used to choose a nine clustered rectangular grid in the SOM analysis and thus if this makes sense at all. A hierarchical clustering using a Pearson correlation matrix may be more suitable or a PCA followed by a qualified k-means or SOM clustering.

**Analysis was performed with the available Affymetrix software. Unfortunately, this software does not have PCA analysis or hierarchical clustering. The SOM clustering performed was first recommended for the time-course analysis during software training. The method includes selection of various parameters, but generally the defaults were used. We have noted this on pp 8-9.**

Moreover, it is unclear as whether the microarray experiments performed comply with MIAME standards.

**The analysis complied with MIAME standards of analysis. The assays were first performed in 2000, before the MIAME standards were created. We’ve worked to fit the existing data into this configuration as much as possible. We have now stated this on p 9.**

The results should be submitted to a public reservoir like Tox-MIAMExpress at EBI or GEO at NCBI.
Submitting all results to the public reservoirs above is possible but requires special permission from the CDC. The database used is available to the public, although on the CDC domain. Below is the link to this policy, as well as the main statement pertaining to public databases.

CIOs may release data without restrictions for public use through the CDC Information Center or other appropriate venues chosen by CDC. Data may also be shared through the CDC/ATSDR Scientific Data Repository, which is managed by CDC’s Epidemiology Program Office. Finally, CIOs may respond to individual requests.

http://www.cdc.gov/od/ads/pol-385.htm#DATACOVEREDBYTHISPOLICY

For DDH gene Strain 1 time point 15 min SLR with RT-PCR gives 1.23 and from the array analysis –0.4. Strain 3 time point 120 min RT-PCR –3.4 and array 0. These measurements do not resemble very well and such large differences and contradicting results cannot be explained by technical differences. In addition, if the microarray results are that difficult to verify the cutoff of SLR 0.6 should maybe rather be 1 which leave the authors with only 22 genes regulated (at only one time point) out of 5.600 which are on the array.

This array format is no longer in use due to upgrades in the Affymetrix system that limits these types of discrepancies. Further, sequence differences between the probes on the array and those used in RTPCR may play a large role in these results. The RTPCR probes were selected for specific regions of the gene in question, while the array probes may not have been. The probes used on the more recent arrays from Affymetrix are more carefully selected. For example, a later study using the new array format (U133A) and a different chemical found increases in a few members of the DDH family of genes (aldo-keto reductases) but also found genes with similar sequences to be decreased or unaltered. The sequence homology found in these genes with altered patterns of expression may not have been picked up by the less stringent early arrays (noted on p 16).

The above comments give me major concerns as to whether real markers have been found and indeed if an exposure profile can be established from the results, which was the aim of the study. To me the different strains should be treated as biological replicates and genes chosen should be consistently regulated across biological replicates. The profiles shown on figure 2 do not give me confidence that the selected genes can be used to address OTQ exposure.

This is a good point to address. There is variability between normal human strains which limits the use of them as biological replicates. This fit with one of the major points of the study was to look at inter-individual variation between strains. We have added comments needed to describe to the reader the reasoning behind the use of these strains as individual experiments in looking for variability in response and showing that this is possible to determine with microarray analysis. Comments regarding this have been added to pp 5, 14.
Minor concerns:

DDH is missing in the abbreviations

It has been added.

Do MEGM media contain any steroids or other compounds that could influence on the action of OTQ.

Not to our knowledge, but mechanistic information is limited for OTQ.

However, we analyzed changes between treated and control cells that were grown in the same media. If there was an effect of the media, it would be consistent across treatments.

The list of genes on the web should include the SLR and quality measures.

I’ve created another table showing all data as it fits in an excel spreadsheet—it contains intensity levels for all genes present on array, for both treated and controls. I find it unnecessary to describe the design of Affymetrix GeneChips and what a log transformation imply.

Previous internal reviewers at NIOSH had problems with the use of signal log ratio and the blind acceptance of data from software packages without more information on how the data was obtained. This part of the manuscript was added to answer those concerns. It may be superfluous to the initiated but gene array studies are capturing the interest of a wide section of the scientific community.

Houlgatte Rémi:
The authors studied gene expression of epithelial cells following exposure to oxythioquinox (OTQ) with DNA microarrays. Their experiment is correctly designed, but there is major discrepancies between RT-PCR and microarray gene expression measurements. This point is not discussed in any section of the manuscript. Is this a consequence of the multitesting? How reproducible are the data? The authors should convince the reader of the accuracy of their data.

A section has been added discussing these discrepancies, related to sequence variability between primers and probes, as well as the potential of other genes to be probed on the array. For example, some probes, though listed as specific to one gene in a family, are able to probe for other family members, including the CYPs. This is one area of the array that has changed with the newer versions of the Affymetrix gene chips. There have been studies looking at comparisons of array formats, and generally studies look to RT-PCR to confirm array data. In these cases, there is little to no correlation between array formats, and in some cases in RT-PCR data and the arrays. As the arrays are made better, so are the results. In this case, we found that many of the probe sets used on the array were not as specific as we would have liked for particular genes, and were not as specific to the gene in question as the primers used in RT-PCR. Genes were first selected for inclusion in this study based on the microarray data, and were not removed if the RT-PCR data was not a perfect match (Comments added to p 16).

Major points:

1) The authors does not address the question of multitesting: How many false-positive (found differentially expressed only by chance) genes should be expected at 5% risk (p <= 0.05) ?
Clearly, in performing assays for 7000+ genes it would be necessary to generate p-values of $<10^{-7}$ to claim statistical significance when applying the Bonferroni correction. The Bayesian school of thought would have no problem with ignoring Bonferroni. However, we used the gene chip as a hypothesis generating tool and RT-PCR as a confirmatory method. The p values shown are produced by the Affymetrix software in relation to significance of the signal found on the array, and in turn comparison of that to other arrays.

2) RT-PCR (table 2) and microarray data (figure 3) provide inconsistent results:
CYP2A13 at 120 min:
Microarray: SLR=0 for all cell-cultures except strain 3 (SLR=0.8)
PCR: increased in all cell-cultures (SLR from 0.62 to 6.01)
DDH at 120 min
Microarray: SLR=0 strain 2 and 3 and increased (SLR>0.5) in strain 1 and 4
PCR: decreased in all measured cell-cultures (SLR < -0.5)
This should be clarified.
This discrepancy is discussed as above.

3) Page 11: “For example, prohibitin expression is found to be up-regulated in all cell strains after OTQ exposure, with similar expression patterns associated with a decrease in cancer incidence.[21]” RT-PCR results in table 2 show that Prohibitin is always decreased in strain 3.
This has been corrected in the text.

Minor points:
4) Page 2: “Gene expression profiles were determined with selection criteria of ±0.6 signal log ratio (SLR) and p value $<= 0.05$.”
This sentence is unclear and should be improved: signal log ratio and p value are defined only in Materials & Methods section, and a profile is the expression level of large number of genes, and not a gene selection.
The gene expression profile included all genes altered by at least a signal log ratio (SLR) of ±0.6 and p value $\leq 0.05$ in three of four cell strains analyzed.

5) Page 9: “From the full list of genes that fit this criteria; select genes were chosen due to their potential role in carcinogenesis, whether by cell cycle control, immune response or other specific functions.”
“select genes” should be replaced by “selected genes”
This alteration was made.

6) Figures 1 and 2 have no vertical scales. Is it linear or log?
Both are linear, though the SOM graph is shown as relative expression, and no scale is given by the analysis software. This is mentioned in the legend, and Figure 1 has been corrected.

7) Even tough I could reach the NAWQA Pesticide National Synthesis Project at the USGS, I could not find reference 1. Is it a valid a reference?
http://ca.water.usgs.gov/cgi-bin/pnsp/pesticide_use_maps_1997.pl?map=W6047 is the correct website. This will be added. This shows the map of OTQ usage for 1997, not 1992. That map is available at http://ca.water.usgs.gov/pnsp/use92/oxythqnx.html. The statement came from the previous website.

8) Figure 2 has a limited interest. It does not highlight the selection strategy, which is still unclear to me. Why is it 3 different curves? Why only 9 clusters?

More data on this is given in the methods section (Comments added p 9).

9) Legend figure 2: “Asterisks indicate a statistically significant variation in expression from the control level as measured by Tukey’s Biweight analysis.”

There is no asterisk in this figure. What threshold for the test?

This was mistakenly put in for Figure 2 and not Figure 3 and 4. This has been corrected.

10) References should be checked; some references are not valid.


This reference (#9) has altered since it was accessed. Information can now be found at http://envirocancer.cornell.edu/default.cfm, specifically at http://envirocancer.cornell.edu/Topics/Pesticide/Pesticide.cfm#Specific.

Reference 5= http://www.cdpr.ca.gov/docs/toxsums/pdfs/410.pdf


Mike Wang:

In this study, the authors conducted a cDNA microarray analysis of global gene expression profiles of cellular responses to a quinoxaline pesticide, oxythioquinox (OTQ). RNA samples were obtained from four normal human mammary epithelial cell strain established from women who underwent reduction mammoplasty through the Cooperative Human Tissue Network. The alterations of gene expression profiles were assessed using self-organizing map (SOM) method after OTQ treatment for 15, 60 and 120 minutes compared with the vehicle control (DMSO) for 120 minutes. The authors reported the expressions of 36 genes were altered in at least three of four cell strains with over ±0.6 signal log₂ ratio (SLR).

I have several major concerns about this study. First, I am concerned about the study design. Why did the authors decide to use mammary epithelial cells? Whereas several previous studies demonstrated that the OTQ had demonstrated hepatotoxicity in laboratory animals. Hepatocytes and mammary epithelial cells are two different and well-differentiated cell types, and it is possible for them to have different expression...
profiles in response to OTQ exposure. Therefore, the authors should justify the use of human mammary epithelial cells in the study of OTQ-induced gene expression profiles.

The use of mammary epithelial cells was in order to look at inter-individual variation in response to various occupational chemicals. These cells are more easily obtained from normal human donors, and allowed the creation of a large tissue/DNA bank. Further, although OTQ exposure has not been related to breast cancer (and to my knowledge never studied in relation to this), pesticide use has been related to breast cancer. While the use of hepatocytes has been the main area of testing in the past, it would not have been possible to look at inter-individual variation in normal cell strains with hepatocytes, which was a main focus of the study. Most hepatic cell lines are immortalized, and would have limited the analysis of inter-individual variation. Comments can be found on p 5, 14.

Second, the sample size of only four cell strains is not large enough to conduct microarray analysis using primary cell lines established from human tissues. Genetic heterogeneity in human population results in tremendous variations of gene expressions. In microarray analysis, the scales of interpersonal variations are comparable or larger than the true signals, which is usually a major problem of successfully applying this technology on human samples. All of the data of this study, including immunoassay of p53 protein levels (Figure 1), microarray analysis (Figure 3) and RT-PCR (Table 2), clearly demonstrate that there are large interpersonal variations among four cell strains at baseline and each time point of OTQ treatment. Without proper adjustments of the interpersonal variations, it is very difficult to distinguish the true OTQ-induced responses from the random variations of expressions within the human populations. In addition, the batch effect of microarray analysis is another major source of artificial variation that can disturb the interpretation of data mining results. In this study, there were fewer details on how the microarray analyses were conducted, such as whether the RNA samples of vehicle control and each time points of each cell strain were analyzed in one batch experiment or in different batches. There is also a lack the details of data scaling and normalization. Thus, with the consideration of larger interpersonal variations of gene expressions and batch effects of microarray technology, the size of only four primary cell strains from human tissues cannot give enough power to allow the authors to assess the OTQ-induced responses.

Microarray analysis for each cell strain was performed in one day. All controls and treatments were performed at the same time for each individual cell strain. Normalization of the data was performed with the Affymetrix software based on a scaling factor (1500) looking at the results of all genes on the array. This information will be added to the paper. These are all good points, but suggest to me that the paper doesn’t stress part of what we wanted to examine was the inter-individual variation. Comments will be revised to stress the importance of this aspect of the study (p 5, 9, 14).

Third, the data analysis is not conducted properly. The authors initially used SOM method to identify a number of patterns of gene expressions at four time points in each cell strain, and then searched for common patterns across cell strains. According to
SOM patterns, genes with a SLR to vehicle control >0.6 (~ larger than 1.5 fold change) in at least two of four cell strains were considered to be significantly changed their expressions in response to OTQ exposure. However, the microarray analyses using human samples are not like those using laboratory animals, where the individual animal can be regarded as replicate samples because of homogenous genetic background. Considering genetic heterogeneity and large interpersonal variations, an increase or decrease of expression in two or more of only four subjects cannot give statistically significant conclusions. With the available vehicle controls, the approach of paired t-tests of comparing each time point to baseline should be used in selecting genes with altered expressions on a larger size of human samples. Moreover, all RNA samples from each cell strain should be analyzed in the same batch of microarray analysis to control the systematic variations.

All RNA from each cell strain was analyzed in the same batch of microarrays at the same time. All microarrays used were from the same lot as well. There was still array variability, but this was done to help control for that. Further, we did not use these samples as biological replicates, but to examine the variation between them and agree that a larger sample size is needed to confirm any results found here. We currently have over 40 tissues banked and ready for analysis if further studies are to be performed for follow-up. This study was originally undertaken to determine if there were any clear-cut differences between the individual cell strains selected, especially related to p53 haplotypes. This was not found to be the case, so further study is not warranted at this time (p 14).

Lastly, there is a concern that the authors conducted RT-PCR using the same cDNA synthesized from microarray analysis, and did not included housekeeping controls. Since there are large variations among reverse transcriptions, RT-PCR using cDNA synthesized from microarray analysis cannot be considered as an independent analysis to confirm the results of microarray analysis. To confirm the results of microarray analysis, RT-PCR should start from RNA samples. This is a valid point, but in general the RT-PCR is performed on the same cDNA to assess the results of the two methods. Housekeeping controls used were both 18S and GAPDH to normalize for discrepancies between samples. This information will be added. Since the analysis was performed in duplicate, any fluctuations in reverse transcription should have shown up as a bad replicate, and would have been re-analyzed (p 9).

There are several minor concerns about this study:

1. In the Results, Real-Time PCR section, authors claimed: “RT-PCR confirmed data found by microarray analysis for the gene listed.” Authors should give more details and explain how to reach this conclusion.

   Comments were added to this section to try to clarify this.

2. Authors didn’t give the details of how and why to conduct RT-PCR at 12 and 24 hours in Materials and Methods section. What about the cell viabilities for each cell strain after 12 and 24 hours OTQ treatment?
There was no difference in toxicity at these later timepoints as compared to earlier analysis, and viability stayed within range of what was found at 2 hrs.

3. The first paragraph of Discussion section didn’t properly discuss the application of similar microarray studies in determining potential hazards in the environmental and the workplace.
   More has been added to the discussion.

4. In the Discussion section for page 13, the authors should indicate which two strains had the minor variant of p53.
   This was added.

5. In the middle of Page 9, sentence “Table 1 contains … and a list of 7 genes decreased in three or more of the four strain …” is not consistent with the list in Table 1 (Only 5 genes were listed).
   This has been corrected.

6. Figure 1 should give the scale of the Fluorescence Intensity for p53 protein expression.
   This has been added.

7. In the legend of Figure 2, asterisks indicate a statistically significant variation in expression form the control level as measured by Tukey’s Biweight analysis. But there is no asterisk in Figure 2.
   This was incorrectly put into the legend for Figure 2 instead of Figure 3. This has been corrected.

8. Information of statistic tests should be given in the legend of Figure 3.
   This has been added.

9. The legends of both Table 1 and 2 indicate: “Genes altered in three of four cell strain in italics.” But there are not italics in both tables.
   This statement has been removed, as both tables have been reformatted.

Yi Yang:
The manuscript by Gwinn et al studied the gene expression profile of normal human mammary epithelial cells following exposure to oxythioquinox. Overall, the paper is well written, but several changes should be made. They are listed below:

1. The study admitted “some inter-individual variability between donors”. It would be desirable to include a heatmap showing the global gene expression profiles across all the donors and time points. Are the expression profiles clustered by donor, or by time points? It is also desirable to include any demographic information for the discussion.
   At this time, we are not able to perform this type of analysis. The Affymetrix software was used for all data analysis, and is only able to perform SOM or k-means
clustering. Future versions of this software will be able to carry out hierarchical analysis as well as possibly heatmap analysis, but this is still being beta tested.

2. Donor 3 had low cell viability (65%) at 120-min time point. Do the authors have any explanations? Was the gene expression profile different from the rest of the treatments? (major)

   **There were no obvious differences with the gene expression profile as compared to the other cell strains. This information was added to the text.**

3. It was not evident in the manuscript to use mammary epithelial cells to monitor OTQ exposure and toxicity. Are there any breast cancer incidents related to OTQ exposure? (major)

   **See comments, reviewer 3. Comments have been added to the text on page 4.**

4. There is a lack of clarifications on time points when describing gene expression changes. For example, were those genes selected in table 1 based on changes at all the three time points, or a single time point? (major)

   **Genes were selected if altered in any time point, because the rate of response may vary between individuals.**

5. Page 7: Need a better description of average percent variability between arrays. (minor)

   **This definition has been added to.**

6. Figure 2 has no scales on the Y-axis. (minor)

   **This has been corrected.**

7. Table 1 and 2 do not have the proper font (italics) that indicated in the legend. (minor)

   **This statement has been removed, as both tables have been reformatted.**