Front. Biol. 2012, 7(1): 65–72
DOI 10.1007/s11515-011-1177-8

REVIEW

Application of microarray technology in *Drosophila* ethanol behavioral research

Awoyemi A. AWOFALA[1,2]

[1] *School of Life Sciences, University of Sussex, Falmer, Brighton BN1 9QG, United Kingdom*
[2] *Department of Biological Sciences, Tai Solarin University of Education, Ijebu-Ode, Ogun State, Nigeria*

Abstract **Gene expression profiling of *Drosophila melanogaster*, an invertebrate model organism, applied to DNA microarray promises to provide novel insights into the important pathways and molecules that may contribute to the risk of alcohol abuse and addiction. Instead of studying one gene at a time, the technology provides a snapshot of transcriptional changes at once, and offers unprecedented opportunities to understand the molecular complexity of alcohol-seeking behavior including addiction and dependence.**

Keywords *Drosophila*, behavior, microarray, gene expression, ethanol, addiction

Introduction

Studies of the neural basis of behavioral responses to alcohol exposure in *Drosophila* continue to play a vital role in understanding the neuronal circuits that mediate alcohol response and possibly addiction and or dependence in humans. The use of this model organism in alcohol behavioral research is in part due to its possession of the following features: 1) fewer ethical boundaries than in human studies, 2) higher efficiency for identifying genes underlying human alcohol-seeking behavior, 3) the ease with which genetic make-up can be determined and or manipulated by the experimenter, 4) possible breeding strategies that cannot be performed in humans, and 5) the relative ease with which variables such as environment, drug-intake, and conditional motivators can be controlled. In addition, it has been recently established that *Drosophila* can be used to model some features of alcohol addiction such as preference (Devineni and Heberlein, 2009) and reward (Kaun et al., 2011). Thus, while certain other key features of alcohol addiction in humans cannot be adequately modeled in *Drosophila*, studying fairly simple behaviors induced by acute ethanol exposure in this invertebrate model not only promises to help gain mechanistic insights into the more complex process of alcohol addiction but also help in advancing our understanding of alcohol-seeking behavior and mental health. To illustrate this, the use of *Drosophila* model has helped to uncover important pathways in the genetic and neural networks mediating ethanol responses in mammals. One of such is the previously documented role of cyclic adenosine 3, 5-monophosphate (cyclic AMP) signaling pathway in the responses to ethanol in *Drosophila*, mice and in human addictive brain (Moore et al., 1998; Wand et al., 2001; Yamamoto et al., 2001).

Over the years, *Drosophila* behavioral geneticists have conducted mutant screenings to identify mutants with altered ethanol-induced behaviors (e.g., sensitivity and tolerance to ethanol) using unbiased screens. Such screens have typically identified mutants at a frequency of 1% or less [e.g., altered ethanol sensitivity, 12 mutants of 5000 screened (Moore et al. 1998) or 23 mutants of nearly 30000 screened (Singh and Heberlein, 2000); altered ethanol tolerance, 2 mutants of 404 screened (Scholz et al., 2005)] (Berger et al., 2008).Thus, the time required to screen such large numbers of genes of which only small fractions were implicated in some forms of ethanol-induced behaviors constituted a practical limitation into the behavioral research on alcohol response and possibly addiction. Nevertheless, recent advances in technology, including high throughput microarrays, have allowed the screening of thousands of genes in a relatively short period of time. The microarray detects more genes and gene networks

Received September 5, 2011; accepted October 10, 2011
Correspondence: Awoyemi A. AWOFALA
E-mail: a.a.awofala@gmail.com

that were later implicated in some aspects of ethanol-induced behaviors (Awofala et al., 2011; Kong et al., 2010; Morozova et al., 2007; Morozova et al., 2006).

To date, a limited number of published studies (Awofala, 2011a, b) have used microarrays to identify candidate genes and relevant metabolic and signaling pathways mediating ethanol response in *Drosophila*. The present review focuses on methods used in microarray technology as applicable to *Drosophila* ethanol behavioral neuroscience research. Specifically, it discusses the different platforms that are available but with bias to oligonucleotide array (see text), the data analyses and higher level data interpretation issues, and present examples from the use of *Drosophila* arrays to examine changes in transcriptional program (i.e. molecular signatures of the state of activity) between ethanol exposed and non-exposed *Drosophila* cells (samples).

Microarray platforms

A microarray is a microscopic array of large sets of DNA sequences immobilised on solid substrates (Eisen and Brown, 1999). Expression profiling experiments usually involve measuring the relative amount of mRNA expressed in two or more experimental conditions. This is based on the assumptions that most mRNAs get translated into proteins and that most changes at the mRNA levels suggest changes at the level of the protein encoded by the mRNAs. The use of gene expression profiling (such as microarrays) provides a snapshot of all the transcriptional changes in a biologic sample. This high throughput method, unlike the traditional techniques such as Southern and Northern blots that focus on a single gene or limited set of genes, can be used to monitor thousands of mRNA transcripts in a cell and facilitates the discovery of totally novel and unexpected functional roles of genes. The power of microarrays have been applied to a range of applications including discovering novel disease subtypes, developing new diagnostics tools and identifying mechanisms of disease or drug response (Slonim and Yanai, 2009).

There are two main technologies for making microarrays: robotic spotting and *in situ* synthesis. Spotted microarray is the technology by which the first microarrays were produced. The array is made using a spotting robot via three main steps: 1) making the DNA probes to put on the array, 2) spotting the DNA onto the glass surface of the array with the spotting robot, and 3) postspotting processing of the glass slide. The DNA probes used in the spotted array can either be cDNA probes or oligonucleotides. cDNA probes are made via highly parallel polymerase chain reaction (PCR) while oligonucleotides probes are presynthesised for use on the arrays (Stekel, 2003).

In situ synthesized oligonucleotide arrays are fundamentally different from spotted arrays: instead of presynthesising oligonucleotides, oligos are built on the glass array one base at a time. At each step, the base is added via covalent reaction between the hydroxyl group 5′ of the terminal base and the phosphate group of the next base. A protective group exists on the 5′ of the base being added and prevents the addition of more than one base at each step. Following addition, there is a deprotection step at which the protective group is converted to a hydroxyl group using either light (i.e. photodeprotection as in the case of Affymetrix® Genechips Technology, Nimblegen and Febit) or acid (i.e. chemical deprotection as in the case of Rosetta, Agilent and Oxford Gene Technology) to allow addition of the next base. Most studies utilizing microarray in *Drosophila* alcohol research have used the Affymetrix Genechips and is the focus of this review.

Experimental design

In a typical *Drosophila* microarray experiment, applied to alcohol research, RNA is extracted from *Drosophila* tissues, labeled with a fluorescent marker (usually a biotin marker), and hybridized to the Affymetrix arrays. The fluorescent intensity values at each probe location are then determined from the scanned optical image of the array, and these intensities reflect the abundance of the targeted RNA in the sample of interest. However, careful experimental design is crucial for the success of such a microarray experiment. Thus, it is important during the design stage to identify all the variables to be compared and to ensure that the proposed design allows their measurement. It is equally important to be aware of variables such as fly age or date of sample collection that might confound the distinction between the compared groups. The use of sufficient replicates is another design issue in microarray; this is because without replicates, it will be practically impossible to estimate statistical significance and reliability of the observed changes in expression and as such result in an increased number of both false positives and false negative errors in detecting differentially expressed genes (Lee et al., 2000). Indeed, a good experimental design should help to remove the often encountered statistical problems of confounding and bias from the microarray data.

In *Drosophila* alcohol research, a case-control blocked design predominates. In such a design, two samples from the same genetic background e.g., ethanol treated and untreated *Drosophila* tissues are compared directly. This is an excellent design as its helps to eliminate the problem of sample variability and genetic heterogeneity. The use of oligonucleotide array in *Drosophila* alcohol research also has its own advantage as the technology has been adjudged to outperform the cDNA array in terms of number of spots per array and the spots' homogeneity (Shi et al., 2006).

Processing methods for Affymetrix data

Microarray data are often coupled with many sources of variations. Irizarry et al. (2003b) explained two sources of variations in high density oligonucleotide arrays-(a)

interesting and (b) obscuring variations. Interesting variations are sources of genetic variation between two experimental conditions, for example, high expression of a particular gene or genes may result from a disease process due to variation between diseased and normal tissue. Obscuring sources of variations are variations (in the form of observed expression levels) introduced during sample preparation, manufacture of the arrays or the processing of the arrays (labeling, hybridization and scanning). Thus, before data from multiple microarray experiment can be pooled into a single analysis the data must first be normalized and corrected for possible sources of obscuring (or technical) variations.

Various methods which include background adjustment, normalization and summarization have been proposed for normalizing Affymetrix GeneChip arrays (Huber et al., 2005). Background adjustment is required to remove the intensity caused by non-specific hybridization and the noise in the optical detection system. Normalization is required for removing experimental variation such as different levels of labeling and account for non-specific hybridization. Essentially, normalization is used to compare intensity data from multiple arrays. Finally, summarization is required to calculate expression levels when transcripts are represented by multiple probes (Huber et al., 2005).

Many algorithms are available for normalizing GeneChip data and also for calculation of their expression values. The most commonly used are RMA (Irizarry et al., 2003a; Irizarry et al., 2003b), GCRMA (Wu et al., 2004), MAS5.0 or its successor GCOS (Affymetrix, 2001a), dChip (Li and Wong, 2001). MAS 5.0 or its successor GCOS is used by Affymetrix systems and makes use of information from only one microarray and also incorporates both the PM and MM probes (Affymetrix, 2001b). In contrast, the model based algorithms involving RMA and dChip incorporate information from multiple microarrays to calculate the expression of a gene by fitting probe response patterns over multiple arrays with a multiplicative model in dChip (Li and Wong, 2001) and an additive model in RMA (Irizarry et al., 2003ab). These fitted models are used to detect abnormal probes, which are subsequently excluded from gene expression calculation (Millenaar, et al., 2006). The GCRMA algorithm makes use of two model types, namely GC and RMA (Wu and Irizarry, 2004) thereby using a method of background adjustment that incorporates the physical model of the GC content of the probe. It should be noted that while MAS5.0 and dChip in the PMMM mode use both PM and MM signals to calculate gene expression, dChip PM mode (Li and Wong, 2001), RMA (Irizarry et al., 2003a; Irizarry et al., 2003b) and GCRMA (Wu and Irizarry, 2004) only use the PM information to calculate gene expression. It should be noted that the MM data in GCRMA is used for modeling the background effect and hence is not entirely discarded (Wu and Irizarry, 2004).

Many studies have been carried out to bench mark the effectiveness of these normalization algorithms (Bolstad et al., 2003; Verhaak et al., 2006; Qin et al., 2006; Millenaar, et al., 2006). However, no definite conclusions have been reached on which of the algorithms are most effective for normalization and calculation of gene expression. This is because these algorithms measure gene expression and normalize the data in different ways, and it appears that the effectiveness of these methods can be influenced by the size and the type of data set. For instance the effect of the four pre-processing strategies involving dChip, RMA, GCRMA and MAS5.0 on expression level measurements, detection of differential expression, cluster analysis, and classification of samples have been assessed (Verhaak et al., 2006). The sample used involved gene-expression data of 285 patients with acute myeloid leukaemia (AML) and 42 samples of tumor tissue of the embryonic central nervous system (CNS). It was found that in most cases, the choice of these algorithms has a relatively small influence on the final analysis outcome of large microarray data (AML data set), but has a more profound effect on the outcome of the small microarray data (CNS data set) (Verhaak, et al., 2006). In addition, another study evaluated the effect of different processing methods on oligonucleotide arrays via quantitative real-time PCR and found no advantage on the choice of one algorithm to the other (Qin et al., 2006).

Previous studies have shown however, that algorithms making use of quantile normalization offer the simplest and quickest normalization methods and also gave the most reproducible results on gene expression and the highest correlation coefficients with Real Time RT-PCR data (Bolstad et al., 2003; Millenaar et al., 2006). Both RMA and GCRMA incorporate the use of quantile normalization by using data from all arrays to create the same empirical distribution of intensities for each array (Irizarry et al., 2003ab). Their expression measures are based on different background correction methods, but the same quantile normalization and expression value summarization using the median polish algorithms (Irizarry et al., 2003a; Irizarry et al., 2003b; Wu and Irizarry, 2004). Quantile normalization in RMA and GCRMA functions is followed by a log-transformation step. In this process, the background corrected and quantile normalized probe values are log-transformed usually to a log base 2 (Irizarry et al., 2003b). This stage is necessary to facilitate easy comparison of data of different orders of magnitude. Summarization in both RMA and GCRMA results in a set of probe level expression values for each array involved in the gene expression analysis. This involves the use of a robust procedure such as median polish to estimate model parameters for correcting for outlier probes (Irizarry et al., 2003ab). Figure 1 shows correlation fold changes between the common RMA and GCRMA's uncovered ethanol responsive *Drosophila* genes.

Statistical analysis

The main goal of any microarray data analysis is to detect

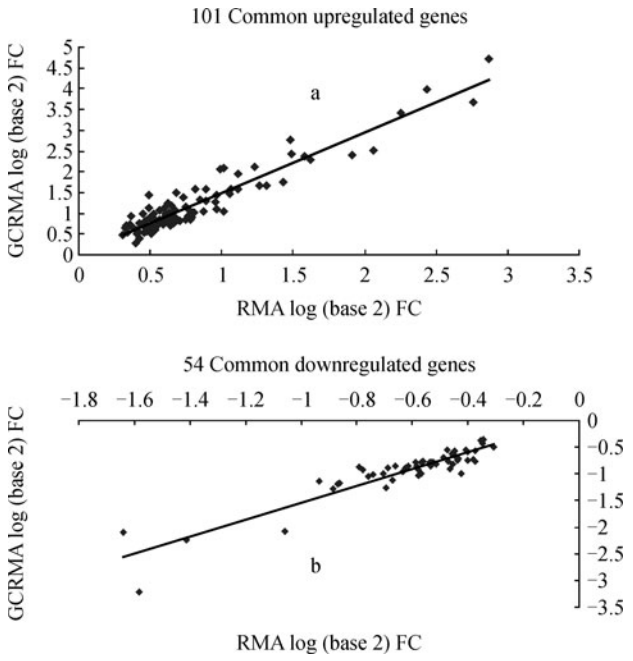

Figure 1 Correlation of fold change between RMA and GCRMA normalization methods. (a) RMA Upregulated genes versus GCRMA upregulated genes. (b) RMA downregulated genes versus GCRMA downregulated genes. The solid lines represent a linear regression fit. Linear fit: RMA upregulated genes versus GCRMA upregulated genes, $y = 1.4594x + 0.0237$, $R^2 = 0.897$; RMA downregulated genes versus GCRMA downregulated genes, $y = 1.584x + 0.0436$, $R^2 = 0.8464$. The common genes were selected based on Benjamini and Hochberg's false discovery rate method (fdr $\leqslant$ 0.05).

differentially expressed genes and a number of statistical tests are available to achieve this. For instance, in a typical study investigating the effect of acute ethanol exposure on gene expression in *Drosophila*, the approach will involve detecting differential gene expression between ethanol-treated and control samples in an Affymetrix *Drosophila* array system. To accomplish this, a null hypothesis of no expression level difference between the two sample conditions is proposed. The alternative hypothesis is that there is difference in the level between the two sample conditions. The hypothesis testing can be carried out by performing a statistical test (in this case, *t*-test) on the expression values of the gene of interest measured in the two conditions. This result in a computed value which can be compared with a threshold t value known as t_α calculated from a t-distribution model and a desired significance level (Cui and Churchill, 2003). The *t*-test assesses whether the means of two groups are statistically different from each other. This statistic achieves this goal by examining the differences between the means relative to the spread or variance of the data (Olson, 2006).

Other statistical tests often used in gene expression analysis include the fold change and the ANOVA. The ANOVA test is used if three or more groups are being considered and compared (Nadon and Shoemaker, 2002). This can either be one-way ANOVA if only one factor is being examined or two-way ANOVA when examining two factors (Olson, 2006). Fold change (FC) is the simplest method for identifying differentially expressed genes and is based on the observed ratio (or average of ratios) between two conditions (Cui and Churchill, 2003). It is the ratio of the measured value for an experimental sample to the value for the control sample. This test is often regarded as an inadequate statistical test because it does not incorporate variance and the differentially expressed genes are not selected based on any significant level of confidence (Miller et al., 2001). Many researchers use FC because it works well for ranking results. This is presumably because all transcripts go through the same processing together, and therefore have similar variances (Allison et al. 2006). Both the FC and the *t*-test statistical criteria can be summarized using an easy-to interpret graph known as volcano plot (Fig. 2). A volcano plot is a device that arranges genes along dimensions of biologic and statistical significance (Cui and Churchill, 2003). Thus, it places genes on a two axis coordinate systems. The y-coordinate corresponding to statistical difference is the negative lg of the *p*-values for the corresponding statistical differences between the two sample conditions. The x-coordinate, corresponding to biologic effects, is the $\log_2$ of the FC between two sample conditions. Genes with statistically significant differential expression (i.e. genes that shows both statistical significance and biologic significance) according to the gene-specific *t*-test will lie above an arbitrarily chosen horizontal threshold line (Cui and Churchill, 2003).

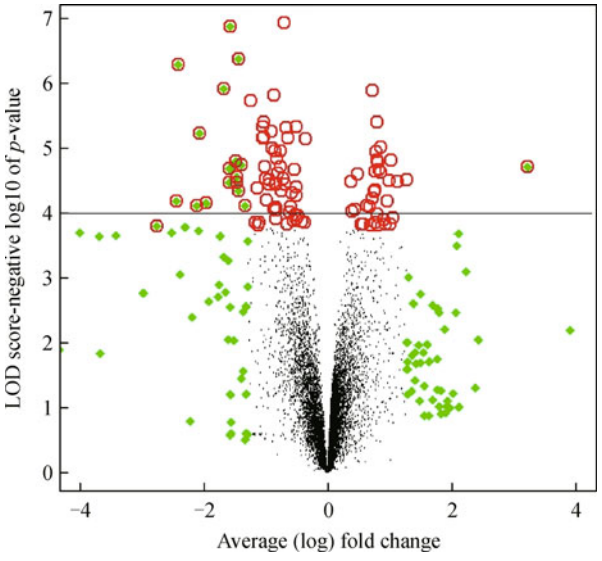

Figure 2 Volcano plot with moderated t-statistics made from microarray data. It shows 100 potentially interesting genes from a biologic standpoint.

However, all statistical inferences are associated with a probability of being incorrect (Nadon and Shoemaker, 2002). Methods are therefore needed to minimise inferential errors

such as type I error (false-positive error), type II error (false-negative error) and the long range error rate (which defines the expected error rate if experiments and analyses of the type under consideration were repeated an infinite number of times) (Allison et al., 2006). Type I error occurs when the null hypothesis is incorrectly rejected. In a microarray experiment an accumulation of type I errors for each gene can result in a substantial number of false positives (genes incorrectly identified as differentially expressed). Conversely, type II error occurs when the null hypothesis is incorrectly accepted resulting in false negatives: an example is failing to identify genes as differentially expressed when they are.

The percentage of the inferential error discussed above can be determined using *p* values. For example, an error rate of 1% (i.e. *p*-value of 0.01) means that on average there will be one false positive for every 100 genes identified as differentially expressed. This might be acceptable in an individual test, but in a microarray experiment with very large number of genes, a considerably high number of false positives results may be found. This therefore calls for an adjustment in multiple hypothesis testing (Benjamini and Hochberg, 1995; Dudoit et al., 2003). Accordingly, two methods have been proposed to address the problem of multiple testing:

Family-wise error-rate control (FWER): Using FWER, the probability of finding at least one false positive is minimised. It is the overall probability that at least one gene is incorrectly identified in the list of differentially expressed genes over a number of statistical tests. For instance, if we identified 1000 genes with an adjusted FWER *p* value of 0.01, then, there is a 1% chance of having one false positive in the list of 1000. The single step Bonferroni correction is the best known method to control the FWER. It defines an effective rate as the standard false positive rate divided by the number of tests conducted (e.g. 0.01/1000) (Nadon and Shoemaker, 2002). This means that every individual gene must have a *p* value lower than 0.00001 to be significant. This highly stringent control often results in an increased rate of false negatives results (Nadon and Shoemaker, 2002). Thus, FWER is more appropriate for analyses in which a single positive is unacceptable, such as comparing various drug treatments with a control. Other known methods of FWER include the step down correction method and permutation based one step correction method (Nadon and Shoemaker, 2002).

False-discovery rate control (FDR): The FDR is the expected proportion of false positives among the rejected hypotheses (Reiner et al., 2003). The FDR adjusts the *p* values so that it reflects the frequency of false positives in a list of differentially expressed genes. Thus, if we identified 1000 genes with an adjusted FDR *p* value of 0.01, then there will be an estimated 10 false positives among the 1000 list of genes. A simple procedure for this approach is that proposed by Benjamini and Hochberg, 1995. FDR is less conservative than FWER and is more applicable in screens for candidate genes in which a small proportion of false positives among the discovered genes is acceptable (Olson, 2006). Hence, the FDR approach has been employed in screening for ethanol regulated genes in *Drosophila* (Awofala, 2011a).

Higher-level data interpretation using gene ontology

Functional annotation: Once a putative list of ethanol-regulated genes has been assembled, some functional analysis is essential to interpret the results. Several programs are available for indentifying pathways, networks or biologic functions that are over-represented in such a given gene list. The functional attributes of each gene fall into three basic categories: 1) molecular function (e.g., actin binding), 2) cellular components (e.g., cytoskeleton), and 3) biologic process (e.g., Toll signaling pathway). Some of these tools include FatiGO (Al-Shahrour et al., 2004), GoMiner (Zeeberg et al., 2003), MAPPFinder (Doniger et al., 2003), MatchMiner (Bussey et al., 2003) and DAVID (Dennis et al., 2003). However, while these tools share many overlapping and related functionalities, a fundamental and widely used tool is the Database for Annotation, Visualization, and Integrated Discovery (DAVID). DAVID has been a program of choice for many researchers working with genomic data, with over 1000 papers citing DAVID from many research institutes world wide (http://david.abcc.ncifcrf.gov/). DAVID analysis can be advantageous because of its very robust nature which stems from its combination of features within a single platform.

Cluster Analysis: Cluster analysis on a list of ethanol-regulated genes and their expression values may lead to valuable insights and important discoveries. Clustering in this sense is often necessary to identify patterns of gene expression and draw meaningful biologic conclusions from them such as predicting functions of gene clusters with previously unknown functions and deducing their causal relationship to the disease under study. Notably, numerous unsupervised (i.e. when the data are organized without the benefit of external classification information) analytic methods have been developed to cluster gene expression data obtained from microarray. Some of these include hierarchical clustering (HC) (Eisen et al., 1998), K-means clustering (Tavazoie et al., 1999) and self-organizing maps (SOM) (Tamayo et al., 1999). Supervised (i.e. the data make use of some external information) analytic methods often used in microarray analysis include k-nearest neighbor classification (Golub et al., 1999), support vector machines (Su et al., 2001), and neural nets (Khan et al., 2001). Ernst and Bar-Joseph (2006) recently proposed a strategy called STEM (short time series expression miner), which is particularly well suited for clustering short time series gene expression data. This approach has been used to identify significant clusters of genes showing time-dependent differential behavior in a study of acute ethanol regulation of gene expression systems in *Drosophila* (Awofala, 2011a).

In a recent time course *Drosophila* acute ethanol microarray study described by Kong and colleagues (Kong et al., 2010), it is often informative to look for general patterns of expression through cluster analysis and thus allows the generation and perhaps validation of the following research questions:

1) Determine what classes of genes show immediate early and late responses respectively and reason whether these behavior could have meaningful biologic interpretations in relation to ethanol response.

2) Determine through cluster analysis whether genes showing time-dependent differential behavior and whose levels are different from base line at the end of the time course are implicated in ethanol tolerance in *Drosophila* when tested experimentally.

3) Determine through experimental validation whether an assumption that genes with similar gene expression patterns and following similar time course show similar response to ethanol sensitivity and tolerance.

4) Finally, determine whether the classes of genes in each cluster are governed by the same regulatory elements or transcriptional factors binding sites. This could be done by analyzing the promoter sequences of early responsive genes to identify the conserved motifs.

Validation of the above hypotheses will offer new directions for further work in conjunction with microarray data. This may offer novel insights into the time-dependent regulation of gene expression monitoring under acute ethanol administration.

Conclusions

Drosophila will continue to play a vital role in research on the neural basis of ethanol-induced behavior. In recent years, the application of microarray has enabled behavioral neuroscientists to analyze global patterns of expression in response to acute ethanol exposure in *Drosophila* (Awofala, 2011a). The use of microarrays in alcohol research may, however, soon be superseded by next generation whole-transcriptome sequencing (Wang et al., 2009). This is particularly useful for assessing low-abundance transcripts that are often missed using microarray technology. However, recent studies have shown that the two transcriptomics technologies are expected to give very similar results (Marioni et al., 2008; Wang et al., 2009). Thus, both the microarrays and the new –omics technologies can now be used to model features of alcohol addiction based on recent studies establishing the use of *Drosophila* in understanding the molecular, genetic and neural mechanisms underlying ethanol preference (Devineni and Heberlein, 2009) and its rewarding properties (Kaun et al., 2011). It should, however, be noted that it may be difficult to formulate a hypothesis about the behavioral role of the genes obtained from microarrays. This is because changes in expression of some of the genes could constitute many alternative events or changes taking place in different tissues or cells or at different times, some of which may be relevant to some ethanol-induced behaviors and not others. Nevertheless, data from microarray experiments can provide powerful information to help determine the molecular pathways underpinning alcohol dependence, the mechanism by which alcohol work, and what gene products may be unique targets for therapy in this disorder.

Acknowledgements

This work was supported by a fellowship from Tai Solarin University of Education. The content is solely the responsibility of the author and does not necessarily reflect the official view of the funding institution.

References

Affymetrix (2001a) Affymetrix Microarray Suite Users Guide, Affymetrix, Santa Clara, CA, version 5.0 edition

Affymetrix (2001b) Statistical Algorithms Reference Guide. Technical report, Affymetrix, Santa Clara, CA

Al-Shahrour F, Díaz-Uriarte R, Dopazo J (2004). FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. Bioinformatics, 20(4): 578–580

Allison D B, Cui X, Page G P, Sabripour M (2006). Microarray data analysis: from disarray to consolidation and consensus. Nat Rev Genet, 7(1): 55–65

Awofala A A (2011a). Acute Ethanol Regulation of Gene Expression Systems in *Drosophila*: A Computational and Behavioral Genetic Approach to Alcohol Addiction. Lambert Academic Publisher (LAP): Germany.

Awofala A A (2011b). Genetic approaches to alcohol addiction: gene expression studies and recent candidates from *Drosophila.* Invert Neurosci, 11(1): 1–7

Awofala A A, Jones S, Davies J A (2011). The heat shock protein 26 gene is required for ethanol tolerance in *Drosophila.* J Exp Neurosci, 5: 31–44

Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: A practical and powerful approach to mMultiple testing. J Roy Stat Soc B Met, 57: 289–300

Berger K H, Kong E C, Dubnau J, Tully T, Moore M S, Heberlein U (2008). Ethanol sensitivity and tolerance in long-term memory mutants of *Drosophila melanogaster.* Alcohol Clin Exp Res, 32(5): 895–908

Bolstad B M, Irizarry R A, Astrand M, Speed T P (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics, 19(2): 185–193

Bussey K J, Kane D, Sunshine M, Narasimhan S, Nishizuka S, Reinhold W C, Zeeberg B, Ajay W, Weinstein J N (2003). MatchMiner: a tool for batch navigation among gene and gene product identifiers. Genome Biol, 4(4): R27

Cui X, Churchill G A (2003). Statistical tests for differential expression in cDNA microarray experiments. Genome Biol, 4(4): 210

Dennis G Jr, Sherman B T, Hosack D A, Yang J, Gao W, Lane H C,

Lempicki R A (2003). DAVID: Database for annotation, visualisation, and integrated discovery. Genome Biol, 4(5): 3

Devineni A V, Heberlein U (2009). Preferential ethanol consumption in *Drosophila* models features of addiction. Curr Biol, 19(24): 2126–2132

Doniger S W, Salomonis N, Dahlquist K D, Vranizan K, Lawlor S C, Conklin B R (2003). MAPPFinder: using gene ontology and GenMAPP to create a global gene-expression profile from microarray data. Genome Biol, 4(1): R7

Dudoit S, Shaffer J P, Block J C (2003). Multiple hypothesis testing in microarray experiments. Stat Sci, 18(1): 71–103

Eisen M B, Brown P O (1999). DNA arrays for analysis of gene expression. Methods Enzymol, 303: 179–205

Eisen M B, Spellman P T, Brown P O, Botstein D (1998). Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA, 95(25): 14863–14868

Ernst J, Bar-Joseph Z (2006). STEM: a tool for the analysis of short time series gene expression data. BMC Bioinformatics, 7(1): 191

Golub T R, Slonim D K, Tamayo P, Huard C, Gaasenbeek M, Mesirov J P, Coller H, Loh M L, Downing J R, Caligiuri M A, Bloomfield C D, Lander E S (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science, 286(5439): 531–537

Huber W, Irizarry R, Gentlemen R (2005). Preprocessing overview. In: Bioinformatics and Computational Biology Solutions using R and Bioconductor, pages 3–12 and 431–442. eds. Gentlemen, R., Carey, V., Huber, W., Irizarry, R. and Dudoit, S. Springer: New York

Irizarry R A, Bolstad B M, Collin F, Cope L M, Hobbs B, Speed T P (2003a). Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res, 31(4): 15e

Irizarry R A, Hobbs B G, Collin F, Beazer-Barclay Y D, Antonellis K J, Scherf U, Speed T P (2003b). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics, 4(2): 249–264

Kaun K R, Azanchi R, Maung Z, Hirsh J, Heberlein U (2011). A Drosophila model for alcohol reward. Nat Neurosci, 14(5): 612–619

Khan J, Wei J S, Ringnér M, Saal L H, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu C R, Peterson C, Meltzer P S (2001). Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nat Med, 7(6): 673–679

Kong E C, Allouche L, Chapot P A, Vranizan K, Moore M S, Heberlein U, Kong E C, Allouche L, Chapot P A, Vranizan K, Moore M S, Heberlein U, Wolf F W (2010). Ethanol-regulated genes that contribute to ethanol sensitivity and rapid tolerance in *Drosophila*. Alcohol Clin Exp Res, 34(2): 302–316

Lee M L, Kuo F C, Whitmore G A, Sklar J (2000). Importance of replication in microarray gene expression studies: Statistical methods and evidence from repetitive cDNA hybridizations. Proc Natl Acad Sci USA, 97: 9834–9839

Li C, Wong W H (2001). Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad Sci USA, 98(1): 31–36

Marioni J C, Mason C E, Mane S M, Stephens M, Gilad Y (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res, 18(9): 1509–1517

Millenaar F F, Okyere J, May S T, van Zanten M, Voesenek L A, Peeters A J (2006). How to decide? Different methods of calculating gene expression from short oligonucleotide array data will give different results. BMC Bioinformatics, 7(1): 137

Miller R A, Galecki A, Shmookler-Reis R J (2001). Interpretation, design, and analysis of gene array expression experiments. J Gerontol A Biol Sci Med Sci, 56(2): B52–B57

Moore M S, DeZazzo J, Luk A Y, Tully T, Singh C M, Heberlein U (1998). Ethanol intoxication in *Drosophila*: Genetic and pharmacological evidence for regulation by the cAMP signaling pathway. Cell, 93(6): 997–1007

Morozova T V, Anholt R R, Mackay T F (2006). Transcriptional response to alcohol exposure in *Drosophila melanogaster*. Genome Biol, 7(10): R95

Morozova T V, Anholt R R, Mackay T F (2007). Phenotypic and transcriptional response to selection for alcohol sensitivity in *Drosophila melanogaster*. Genome Biol, 8(10): R231

Nadon R, Shoemaker J (2002). Statistical issues with microarrays: processing and analysis. Trends Genet, 18(5): 265–271

Olson N E (2006). The microarray data analysis process: from raw data to biological significance. NeuroRx, 3(3): 373–383

Qin L X, Beyer R P, Hudson F N, Linford N J, Morris D E, Kerr K F (2006). Evaluation of methods for oligonucleotide array data via quantitative real-time PCR. BMC Bioinformatics, 7(1): 23

Reiner A, Yekutieli D, Benjamini Y (2003). Identifying differentially expressed genes using false discovery rate controlling procedures. Bioinformatics, 19(3): 368–375

Scholz H, Franz M, Heberlein U (2005). The *hangover* gene defines a stress pathway required for ethanol tolerance development. Nature, 436(7052): 845–847

Shi L, Reid L H, Jones W D, Shippy R, Warrington J A, Baker S C, Collins P J, de Longueville F, Kawasaki E S, Lee K Y, Luo Y, Sun Y A, Willey J C, Setterquist R A, Fischer G M, Tong W, Dragan Y P, Dix D J, Frueh F W, Goodsaid F M, Herman D, Jensen R V, Johnson C D, Lobenhofer E K, Puri R K, Schrf U, Thierry-Mieg J, Wang C, Wilson M, Wolber P K, Zhang L, Amur S, Bao W, Barbacioru C C, Lucas A B, Bertholet V, Boysen C, Bromley B, Brown D, Brunner A, Canales R, Cao X M, Cebula T A, Chen J J, Cheng J, Chu T M, Chudin E, Corson J, Corton J C, Croner L J, Davies C, Davison T S, Delenstarr G, Deng X, Dorris D, Eklund A C, Fan X H, Fang H, Fulmer-Smentek S, Fuscoe J C, Gallagher K, Ge W, Guo L, Guo X, Hager J, Haje P K, Han J, Han T, Harbottle H C, Harris S C, Hatchwell E, Hauser C A, Hester S, Hong H, Hurban P, Jackson S A, Ji H, Knight C R, Kuo W P, LeClerc J E, Levy S, Li Q Z, Liu C, Liu Y, Lombardi M J, Ma Y, Magnuson S R, Maqsodi B, McDaniel T, Mei N, Myklebost O, Ning B, Novoradovskaya N, Orr M S, Osborn T W, Papallo A, Patterson T A, Perkins R G, Peters E H, Peterson R, Philips K L, Pine P S, Pusztai L, Qian F, Ren H, Rosen M, Rosenzweig B A, Samaha R R, Schena M, Schroth G P, Shchegrova S, Smith D D, Staedtler F, Su Z, Sun H, Szallasi Z, Tezak Z, Thierry-Mieg D, Thompson K L, Tikhonova I, Turpaz Y, Vallanat B, Van C, Walker S J, Wang S J, Wang Y, Wolfinger R, Wong A, Wu J, Xiao C, Xie Q, Xu J, Yang W, Zhang L, Zhong S, Zong Y, Slikker W Jr, MAQC Consortium (2006). The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nat Biotechnol, 24(9): 1151–1161

Singh C M, Heberlein U (2000). Genetic control of acute ethanol-induced behaviors in *Drosophila*. Alcohol Clin Exp Res, 24(8): 1127–1136

Slonim D K, Yanai I (2009). Getting started in gene expression microarray analysis. PLOS Comput Biol, 5(10): e1000543

Stekel D (2003). Microarray Bioinformatics. Cambridge University Press: Cambridge

Su A l, Welsh J B, Sapinoso L M, Kern S G, Dimitrov P, Lapp H (2001). Molecular classification of human carcinomas by use of gene expression signatures. Cancer Res, 61:7388–93

Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander E S, Golub T R (1999). Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc Natl Acad Sci USA, 96(6): 2907–2912

Tavazoie S, Hughes J D, Campbell M J, Cho R J, Church G M (1999). Systematic determination of genetic network architecture. Nat Genet, 22(3): 281–285

Verhaak R G, Staal F J, Valk P J, Lowenberg B, Reinders M J, de Ridder D (2006). The effect of oligonucleotide microarray data pre-processing on the analysis of patient-cohort studies. BMC Bioinformatics, 7(1): 105

Wand G, Levine M, Zweifel L, Schwindinger W, Abel T (2001). The cAMP-protein kinase a signal transduction pathway modulates ethanol consumption and sedative effects of ethanol. J Neurosci, 21: 5297–5303

Wang Z, Gerstein M, Snyder M (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet, 10(1): 57–63

Wu Z, Irizarry A R, Gentleman R, Martinez-Murillo F, Spencer F (2004). A model-based background adjustment for oligonucleotide expression arrays. JASA, 99: 909–917

Wu Z, Irizarry R A (2004). Preprocessing of oligonucleotide array data. Nat Biotechnol, 22(6): 656–658, author reply 658

Yamamoto M, Pohli S, Durany N, Ozawa H, Saito T, Boissl K W, Zöchling R, Riederer P, Böning J, Götz, M E (2001). Increased levels of calcium-sensitive adenylyl cyclase subtypes in the limbic system of alcoholics: evidence for a specific role of cAMP signaling in the human addictive brain. Brain Res, 895: 233–237

Zeeberg B R, Feng W, Wang G, Wang M D, Fojo A T, Sunshine M, Narasimhan S, Kane D W, Reinhold W C, Lababidi S, Bussey K J, Riss J, Barrett J C, Weinstein J N (2003). GoMiner: a resource for biological interpretation of genomic and proteomic data. Genome Biol, 4(4): R28