MAIZE GENE EXPRESSION UV RESPONSE PATTERNS REVEAL COORDINATE REGULATION OF MANY GENES

Carletha R. Blanding

A Thesis Submitted to the
University of North Carolina Wilmington in Partial Fulfillment
Of the Requirements for the Degree of
Master of Science

Department of Biology and Marine Biology

University of North Carolina Wilmington

2005

Approved by

Advisory Committee

_____ _____

_____ _____
Chair

Accepted by

_____
Dean, Graduate School

TABLE OF CONTENTS

ABSTRACT

Understanding the mechanisms plants use for acclimation to ultraviolet radiation stress is key in predicting plant responses to our changing environment. Maize inbred line B73 plants were grown in sterilized vermiculite in a greenhouse under low fluence white light for 16 h/day until the majority of the plants were at the three-leaf stage. Plants were then subdivided into five groups of approximately 20 plants for treatments of one, two, four, eight, or twelve hours of UV-B irradiation at a dose rate of 0.024 W m^{-2}. A sixth, control group was placed the same distance under an identical pair of UV bulbs that were covered with Mylar to block all UV-B radiation. Using transcripts from pools of leaf tissue, cDNA microarrays were used to monitor UV-induced alterations in gene expression to gain a better understanding of global gene expression in response to UV-B. Array data were standardized using a variance stabilizing procedure accounting for the many sources of systematic variation that affect measured gene expression levels. Examination of the effects after irradiation at different times indicated that the largest transcriptome response was evident after 12 hr of UV exposure followed by 8 hr, 2 hr, 1 hr, and 4 hr It was also determined that different suites of genes are expressed after specific UV treatments. Overall, it was determined that after 2 hr of UV exposure, rapid transcriptome responses occurred followed by a decrease in gene expression after 4 hr of UV exposure. After 8 hr of UV an increase in gene expression occurred that appeared to level off after 12 hr of UV exposure. With the analysis of gene expression over time, we have been able to identify new gene regulatory patterns in response to UV.

ACKNOWLEDGEMENTS

First, I would like to acknowledge Rebecca Mayo, the one high school teacher responsible for my interest in Biology. I would like to convey my everlasting appreciation and gratitude to my advisor Dr. Ann E. Stapleton for being there and always pushing me to do better and only accepting the best. I would also like to thank her for guidance, and training both of which are essential to my future career. I would like to thank Dr. Susan Simmons for making this collaboration a reality. I thank her for taking the time to have weekly meetings, dedication, advice, support, and concern about me personally and as a student. I want to also express my thanks to the other members of my committee, Dr. Steve Kinsey and Dr. Thomas Shafer, for their expert advice and support. Special thanks to Dr. James Blum and Dr. Dargan Frierson for their endless patience and solutions to my many statistical and programming issues.

Many, many thanks to Yibing Fu for his encouraging words, continuous support, phone calls, friendship, and for taking time to review my thesis. I would like to also thank my families for their support during this journey.

DEDICATION

This thesis I dedicate to my husband and friend Jeffrey W. Blanding and our daughter Erin C. Blanding. They have been there for me through the long days, late nights, homework, exams, summer-long internships, and conferences.

I would also like to dedicate this thesis to the late, devoted Douglas Johnson for making this dream a reality, Chancellor James R. Leutze for taking time out of his busy schedule to listen to my story, Dr. Denis Carter for counseling me during a desperate time of need, and Ralena Wicker for using her wisdom to make this all a possibility.

Lastly this is dedicated to all of those who no one believed in, but with just one opportunity proved that through God the inconceivable is achievable.

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

An intricate relationship exists between an organism's transcriptome and the environmental stresses to which it is exposed. Because the transcriptome is the full complement of activated genes, mRNAs, or transcripts in a particular tissue at a particular time, the slightest change in the environment can provide cues to the mechanisms the organism uses when responding to environmental stresses (Russell, 2003).

As ozone levels continue to decrease, increased amounts of ultraviolet radiation (UV) will reach the earth's surface, adversely affecting plants and animals (Sinha & Hader, 2002). This recent stratospheric ozone depletion has brought about interest in the mechanisms plants use to respond to UV. Ultraviolet radiation is emitted from the sun and is subdivided into three classes; UV-A (400-320 nm), UV-B (320-280 nm) and UV-C (280-250 nm) (Lumsden, 1997). UV-A wavelengths may reach the surface of the earth whereas UV-C wavelengths are absorbed by ozone in the stratosphere and as a result are not present on the earth's surface (Caldwell *et al*., 1989). In contrast, UV-B wavelengths reach the earth's surface and are different from other wavelengths of light because of their high energy and the wide range of biological molecules that absorb them (Jordan, 2002). UV-B is the most biologically effective radiation within sunlight and has been found to be responsible for variation or damage to many plant processes (Tyrell, 1993).

Stapleton (1992) classified this plant damage into two categories: damage to physiological processes or DNA damage. In maize, physiological damage such as leaf rolling, biomass reduction, and fluctuating leaf asymmetry were found in certain inbred varieties (Cartwright *et al*., 2001). UV-B radiation has also been found to significantly

affect the expression of genes in several different biological processes in plants. UV-B causes many different responses in plants and these responses depend on the perception of UV-B radiation, signal transduction pathways, and alterations of gene expression (Jordan, 2002). When plants are exposed to UV-B, alterations in gene expression occur on a large scale. This large-scale modification in gene expression occurs as a result of the plant's cellular response to an environmental stimulus.

In several plants, UV-B has been found to affect genes associated with many protective pathways such as the production of protective flavonoid pigments and DNA repair (Brosche & Strid, 2003; Stapleton & Walbot, 1994). In *Arabidopsis thaliana*, several defense (Douglas *et al*., 1991) and photosynthetic genes (Brosche *et al*., 2002) changed expression levels in response to UV-B irradiation. In maize, the catalase 1, 2, and 3 genes had increased expression levels in response to UV-B (Polidoros & Scandalios, 1997). To date, little is known about the mechanisms plants utilize for acclimation to UV stress. This is in part due to the limitations imposed by the study of few genes at once using techniques such as Northern Blot analysis. With the advent of microarrays, scientists have been able to study thousands of genes simultaneously (Zhu, 2003; Zimselmeier *et al*., 2002) using hybridization to many genes on a single platform to monitor transcriptomes (Schena *et al*., 1998).

Numerous studies have illustrated the power of using microarrays to better understand many biological processes. Using several different maize lines with varying flavonoid content, expression microarrays of maize leaf tissue were used to monitor gene expression in response to four different UV treatments (Casati & Walbot, 2003). These authors found that 14.2% of the genes on the array were regulated by UV, and maize

lines lacking absorbing pigments experienced more dramatic responses to UV than lines with pigments (Casati & Walbot, 2003). This was an observational study that identified up and down-regulated transcripts in response to UV in the field and compared the results to up and down-regulated transcripts in response to UV in the greenhouse. Microarrays were also used to monitor gene expression in different maize organs in response to different levels of UV exposure intensity and duration (Casati & Walbot, 2004). These authors determined by presence and absence tallies, that rapid changes in the transcriptome occurred as early as after 1 hr of UV exposure. (Casati & Walbot, 2004). This study monitored changes in the transcriptome in different maize organs at multiple UV treatments without the use of robust statistics as well. Another study using a single, unreplicated microarray to monitor the transcriptome found that UV-B radiation altered transcript levels of defense genes and photosynthetic genes in *Arabidopsis thaliana* (Brosche *et al*., 2003).

Bar-Joseph (2004) recommends that three factors be considered to ensure reproducibility and effectiveness of a microarray study: experimental design, data analysis, and pattern recognition. Current microarray data do not provide a statistically rigorous overall depiction of how the transcriptome responds to ultraviolet radiation over time or with increasing dose of UV. Thus, using transcripts from pools of leaf tissue, cDNA microarrays were used to monitor UV-induced alterations in gene expression after five different UV treatments. The microarrays used in this present study comprised approximately one third of the maize genome as a conservative estimate (http://www.maizegdb.org/documentation/mgdp/microarray/index.php).

Modern exploratory statistical techniques were used in this study to analyze expression data. Gene ontologies, publicly available "structured, controlled, vocabularies" were used to classify annotated genes in one of three biological domains: biological function, molecular function, and cellular location (Ashburner & Lewis, 2002). This structured framework provided vocabularies regarding genes and their specific functions that were useful for the classification of genes identified in the analysis as significantly differentially expressed. In addition to analysis of gene ontology information, gene description terms were analyzed as well. Using these exploratory statistical techniques, this study aimed to gain a better understanding of global gene expression in response to UV-B and the underlying mechanisms involved with UV-B stress response that are so poorly understood.

MATERIALS AND METHODS

Maize Lines, Growth Conditions and Ultraviolet Treatment

Inbred line B73 was kindly supplied by M. Lee (Lee *et al*., 2002). Individual pots were filled with sterilized vermiculite and seeds were placed one per pot in flats at a density of 36 pots per flat. Flats were placed in the greenhouse at the University of North Carolina Wilmington under low fluence white light for 16 hr /day until the majority of the plants were at the three-leaf stage (8 days).

Plants were divided into control and experimental groups and then subdivided into five groups of approximately 20 plants for treatments of one, two, four, eight, or twelve hours of UV-B irradiation. The UV313 bulbs were suspended approximately 30 cm above the plants. The treatment plants received a total UV dose of 86 Jm^{-2}. Control plants were placed under UV313 bulbs that were covered with Mylar (MylarD, US Plastics, Lima, OH, USA) to block all UV-B radiation. Plants were irradiated centered on solar noon to maximize the amount of sunlight reaching the plants. Immediately following irradiation, bulbs were turned off, and both control and experimental groups were allowed to recover for 3 hr in the greenhouse. The 1 hr group was irradiated from noon to one PM and harvested at four PM. The 2 hr group was irradiated from eleven AM to one PM and harvested at four PM. The 4 hr group was irradiated from ten AM to two PM and harvested at five PM. The 8 hr group was irradiated from nine AM to five PM and harvested at eight PM. The 12 hr group was irradiated from seven AM to seven PM and harvested at ten PM. Following recovery, second and third seedling leaves of each of the control and experimental plants were harvested and immediately frozen in liquid nitrogen.

RNA Preparation, mRNA Purification and Probe Synthesis

Total RNA was extracted from frozen tissue using Trizol (Invitrogen Co., Carlsbad, CA). Pooling, which is recommended for microarrays, was carried out to assure adequate RNA and to decrease biological variability (Peng *et al*., 2003; as cited by Blum *et al*., 2004). After extraction, RNA was shipped on dry ice to Stanford University Department of Plant Biology. Poly$(A)^+$ RNA was isolated using Oligotex (Qiagen Inc., Valencia, CA) and 4 μg of poly $(A)^+$ RNA was used for each cDNA synthesis using Superscript II reverse transcriptase (Life Technologies). cDNA was labeled using 100 μM Cy5-dUTP or Cy3-dUTP (Amersham Pharmacia Biotech, Piscataway, NJ). Excess nucleotides and primers were removed using QIAquick PCR Purification Kit (Qiagen, Valencia, CA).

Hybridization, Scanning, Gridding

Twenty maize cDNA Unigene 01-01-04 slides were used (containing all cDNA sequences available at the time). Array hybridization was carried out using two samples per slide and the dye-switch method (Liang *et al*., 2003); one sample was labeled with each dye, these samples were mixed and then hybridized to a microarray for 15 hr at 60°C. The slides were washed in three wash steps: 2x SSC, 0.5% SDS; 0.5x SSC; and 0.05x SSC. The three washes were at room temperature for 5 min each with gentle shaking.

The slides were scanned with a GenePix 4000B Scanner (Axon Instruments Inc., Union City, CA). Normalization between the Cy3 and Cy5 fluorescent dye emission channels was achieved by adjusting the levels of both image intensities. TIGR Spotfinder, version 2.2.3 was used to perform gridding and quantification of the image files using the

default settings except for the specification to keep raw intensities (Saeed *et al*., 2003).

Statistical Analysis

Data generated from TIGR Spotfinder was imported in SAS Statistical Programming Software (See Appendix A) and formatted. The data was then arsinh standardized (Huber, 2004) with print-tip dependence on location and scale (Yang *et al*., 2002) using vsn in Bioconductor (Gentleman *et al*., 2004) in the R software package (R Development Core Team, 2004) (See Appendix B). Missing intensities were imputed by replacing them with intensities of a replicate within the same timepoint and labeling scheme.

MvA plots were created to ensure effectiveness of standardization procedure (See Appendix C). Array 6 (control labeled Cy3 and treatment labeled Cy5) was removed from the analysis due to an unacceptable MvA plot. Dye duplicate scatterplots for both red and green dyes were created for all microarrays to examine uniformity of labeling across all treatments (See Appendix D).

Standardized data were imported into the SAS Statistical Programming Software Package v8. Master Unigene 1 Large and Master Unigene 1 Small documentation files were downloaded from http://www.zmdb.iastate.edu/zmdb/microarray/arrays.php . For each timepoint separately, intensities were modeled to determine if there was a significant difference between the control and treatment intensities, using the generalized linear model, Proc Genmod (See Appendix A). All p-values were output into tables with corresponding gene names. Calculated p-values were converted into corresponding q-values by the software package Q. Q is a statistical software tool that is founded on the concept of the False Discovery Rate (FDR). Similar to the p-value, which is synonymous to the false positive rate, the q-value provides each gene its own measure of significance

(Storey & Tibshirani, 2003). The difference between the two measures of significance is the false positive rate which is the rate that truly null features are termed significant and the false discovery rate is the rate that significant features are truly null. The False Discovery Rate provides a direct measure of significance by testing each hypothesis separately while taking into account the testing of thousands of hypotheses simultaneously. The program was run using default settings except specifying a Single No. of '.01' and the use of the 'robust' method. Q-values were output into a table with corresponding genes, and genes with q-values less than or equal to .01 (1% FDR) were selected for further analysis.

Standardized treatment intensities of were further analyzed using the ORIOGEN profile analysis software package (Peddada *et al*., 2003). ORIOGEN is a profile analysis program that clusters genes in profiles by creating order-restricted point estimates and using a bootstrap algorithm. Gene intensities were analyzed with ORIOGEN using the following parameters. The total number of dose groups or timepoints was 5, the vector of sample sizes per dose/timepoint was 4, 3,4,4,4, there were 100,000 bootstrap samples run, and a p-value of .001 was applied. All profiles were selected for analysis except the non-cyclic profile.

In order to analyze the ontology information of significant genes, a Perl code was written to extract all known gene ontology identification numbers from the TIGR gene indices maize database located at www.tigr.org. Approximately 33% of the genes on the array were annotated in TIGR. Gene ontology ID numbers (GOID) were merged with significant gene lists from each timepoint and treatment profile lists for further analysis. To determine if significant gene lists and profile lists contained GOID's that were

over/under-represented, a chi-square test for independence with one degree of freedom was implemented in Microsoft Excel (see Appendix D for R formatting code). Each individual list of GOID's was compared to the entire gene list to obtain observed and expected frequencies for chi-square analysis. Generated p-values were analyzed in Q to control for false discovery rates and GOID's with q-values less than or equal to '.01' were considered to be over/under-represented within each specific list.

ALACK, an automated lexical analysis program was utilized to assign p-values to over-represented gene description terms present in specific lists in comparison to the master list, which in this case was the list of all description terms available for genes on the microarray (Kim & Falkow, 2003). This method has proven to work effectively and was utilized to aide in the identification of new functional classes of UV-regulated genes (Blum *et al*., 2004). We used ALACK to analyze each significant gene list, each mutually exclusive significant gene list, and each treatment profile list. Over-represented gene description terms with p-values equal to or exceeding .90 were reported in Tables 3-8.

RESULTS

Result of Data Standardization

Maize seedlings were grown to the 3-leaf stage under conditions similar to the amount of ultraviolet radiation exhibited in nature. Groups of plants were exposed to one, two, four, eight, or twelve hours of ultraviolet radiation. Data standardization was carried out using an arsinh print-tip dependent procedure. To ensure the effectiveness of the data standardization, MvA plots were created for each microarray (Figures 1-20). Upon comparison of all MvA plots, microarray 6 was removed from subsequent analysis due to its nonconformity to the standardization procedure exhibited by its non-uniform and scattered appearance (Figure 6). Also, because it is known that dye bias exists and can cause unwanted variation in data analysis (Liang *et al*., 2003), dye duplicate scatterplots for both red and green dyes were created for all microarrays to examine uniformity of labeling across all treatments (Figures 21-40). There were no irregular patterns of variation in labeling detected, so all samples were further analyzed.

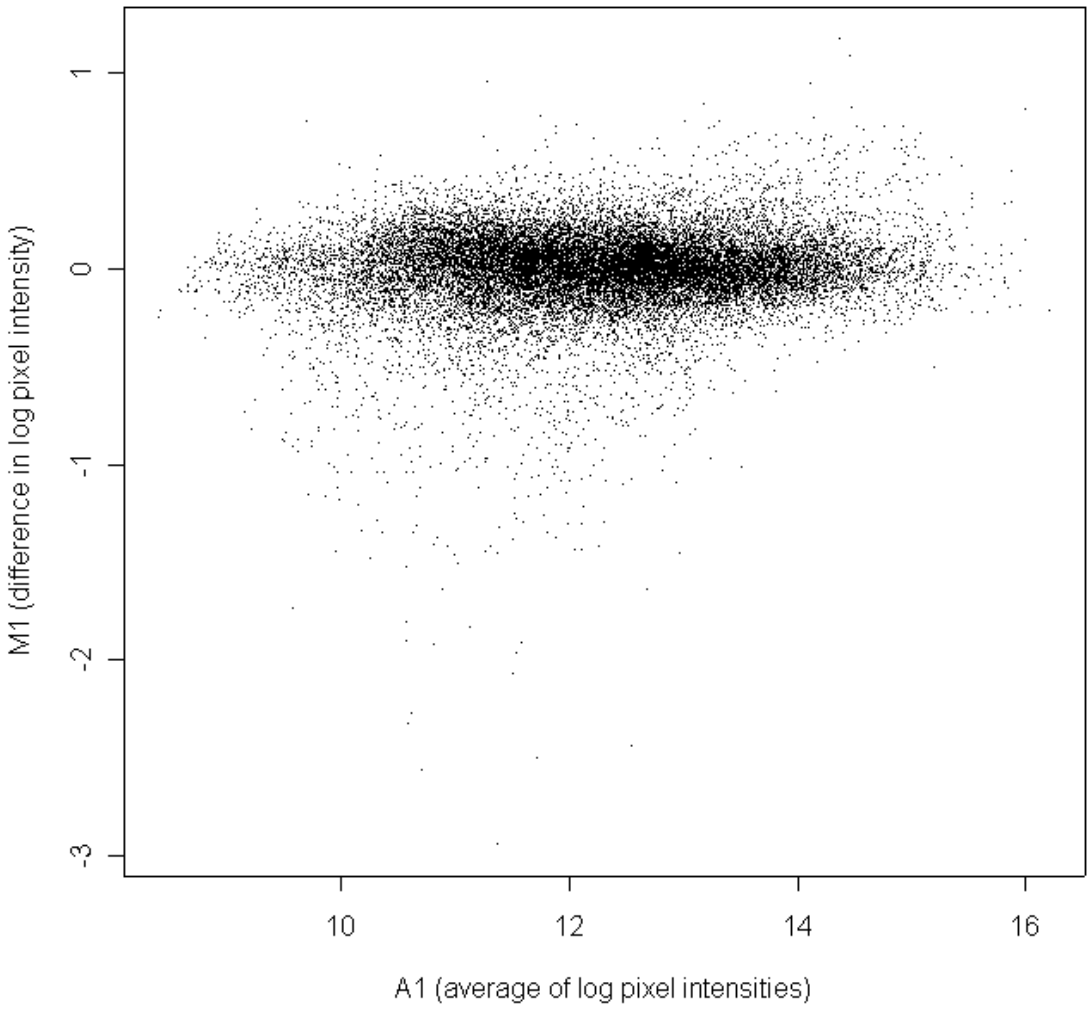

Figure 1. MvA Scatterplot for Microarray 1.
Where M = LogR-LogG, and A = ((LogR+LogG)/2). Log R equals the logarithmic value of the red intensity and Log G equals the logarithmic value of the green intensity. Pixel intensity equals the amount of signal from gene hybridization.

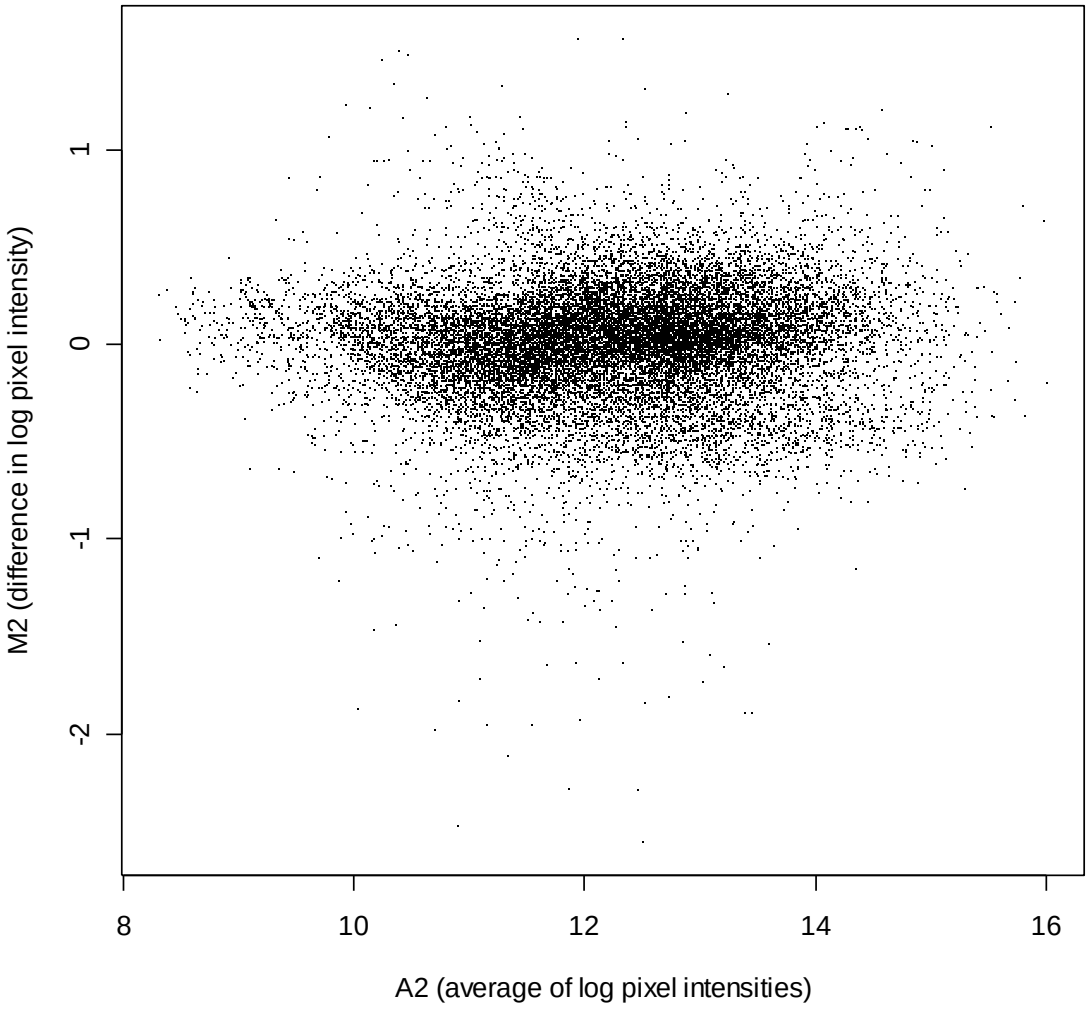

Figure 2. MvA Scatterplot for Microarray 2.
Where M = LogR-LogG, and A = ((LogR+LogG)/2). Log R equals the logarithmic value of the red intensity and Log G equals the logarithmic value of the green intensity. Pixel intensity equals the amount of signal from gene hybridization.

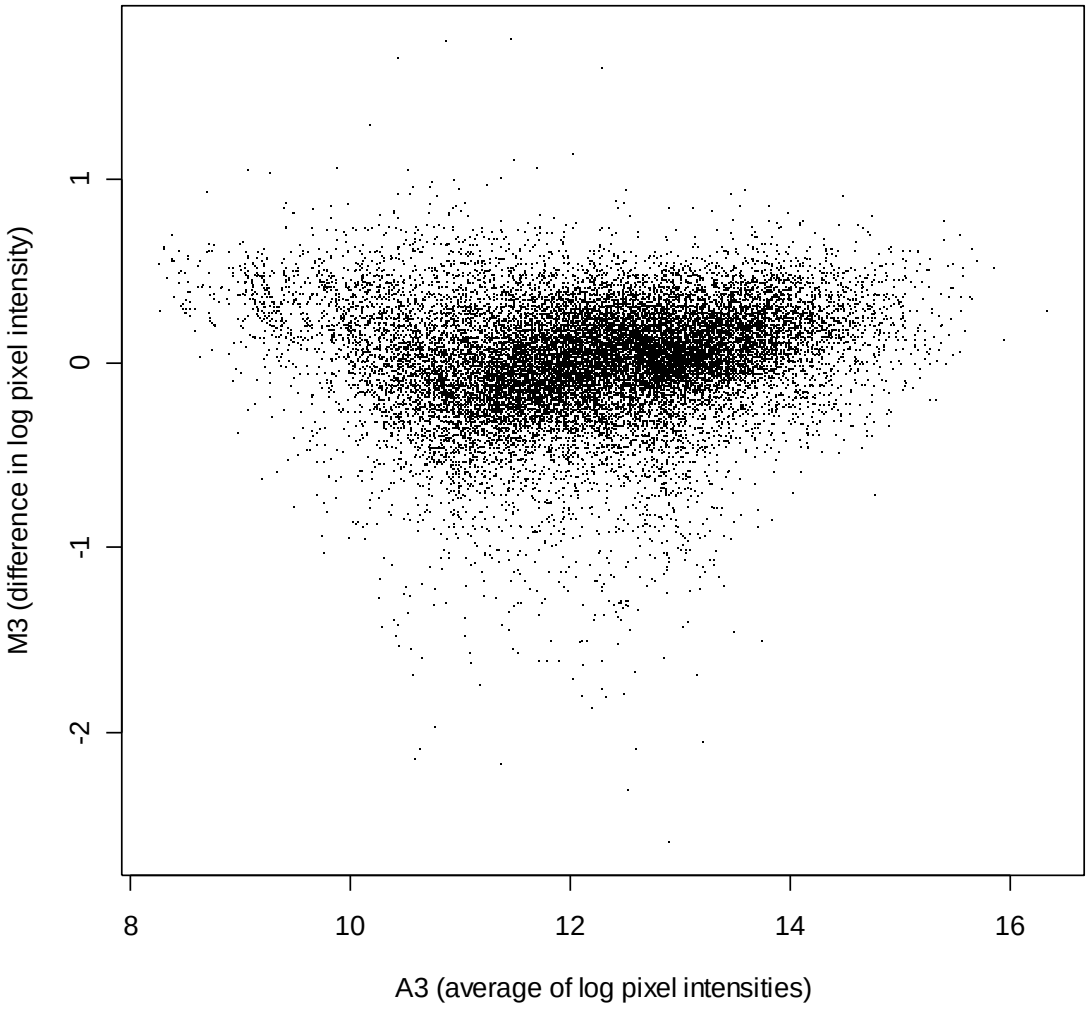

Figure 3. MvA Scatterplot for Microarray 3.
Where M = LogR-LogG, and A = ((LogR+LogG)/2). Log R equals the logarithmic value of the red intensity and Log G equals the logarithmic value of the green intensity. Pixel intensity equals the amount of signal from gene hybridization.

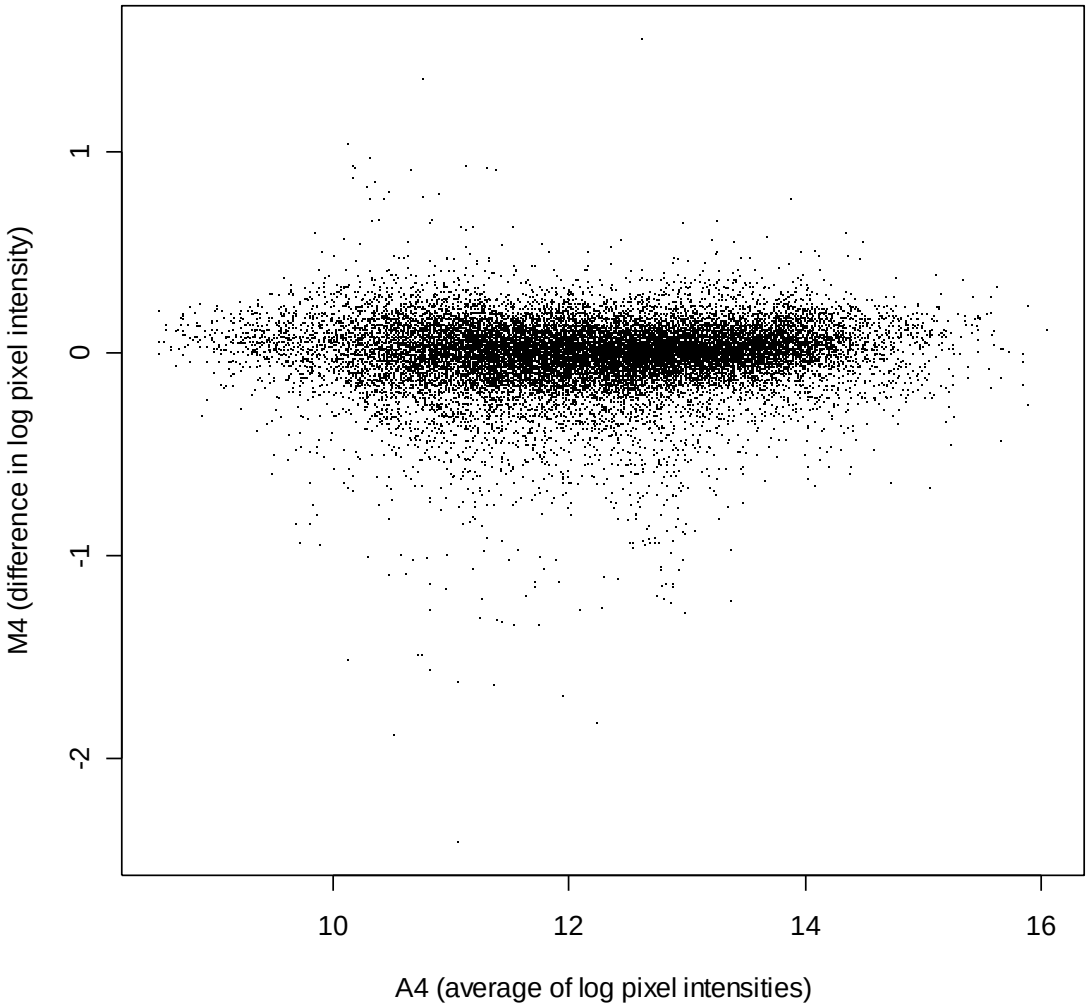

Figure 4. MvA Scatterplot for Microarray 4.
Where M = LogR-LogG, and A = ((LogR+LogG)/2). Log R equals the logarithmic value of the red intensity and Log G equals the logarithmic value of the green intensity. Pixel intensity equals the amount of signal from gene hybridization.

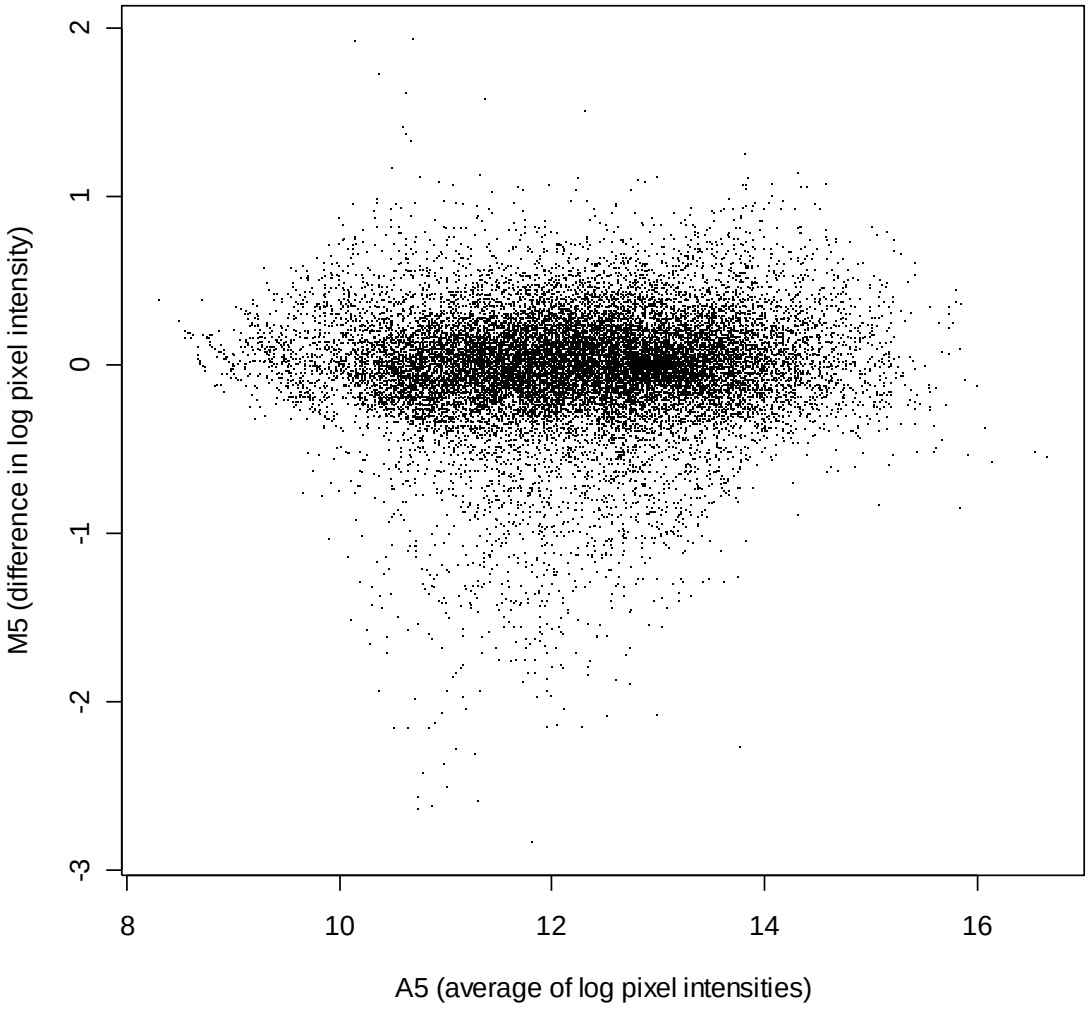

Figure 5. MvA Scatterplot for Microarray 5.
Where M = LogR-LogG, and A = ((LogR+LogG)/2). Log R equals the logarithmic value of the red intensity and Log G equals the logarithmic value of the green intensity. Pixel intensity equals the amount of signal from gene hybridization.

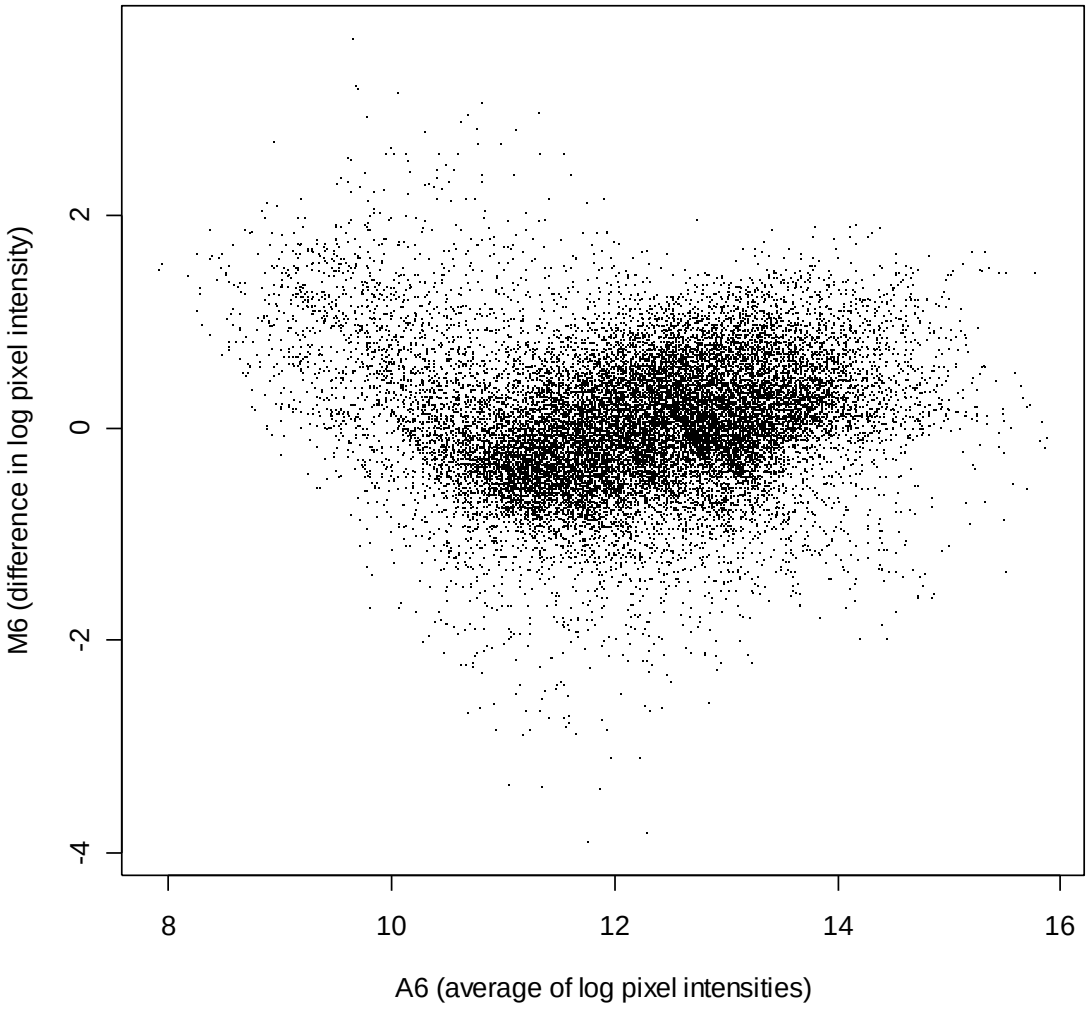

Figure 6. MvA Scatterplot for Microarray 6.
Where M = LogR-LogG, and A = ((LogR+LogG)/2). Log R equals the logarithmic value of the red intensity and Log G equals the logarithmic value of the green intensity. Pixel intensity equals the amount of signal from gene hybridization.

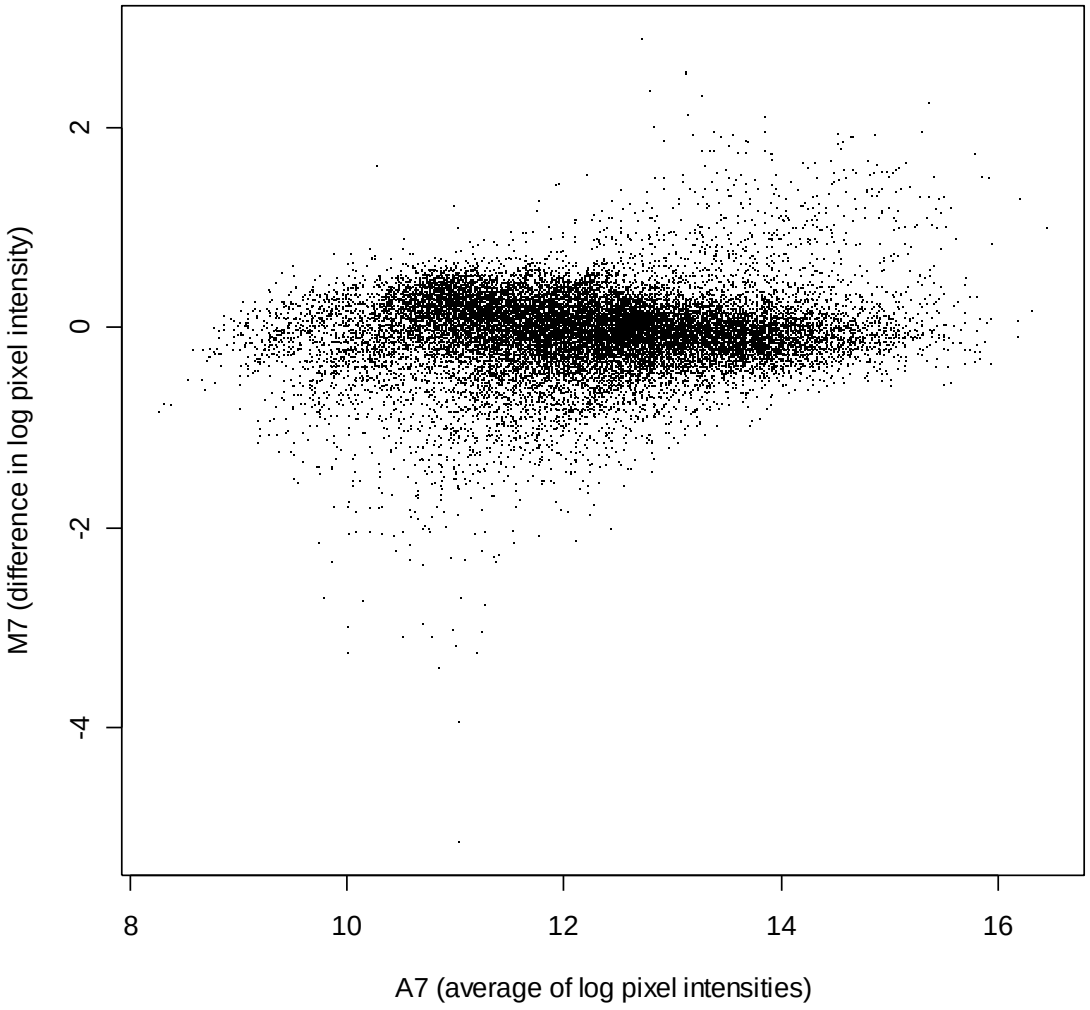

Figure 7. MvA Scatterplot for Microarray 7.
Where M = LogR-LogG, and A = ((LogR+LogG)/2). Log R equals the logarithmic value of the red intensity and Log G equals the logarithmic value of the green intensity. Pixel intensity equals the amount of signal from gene hybridization.

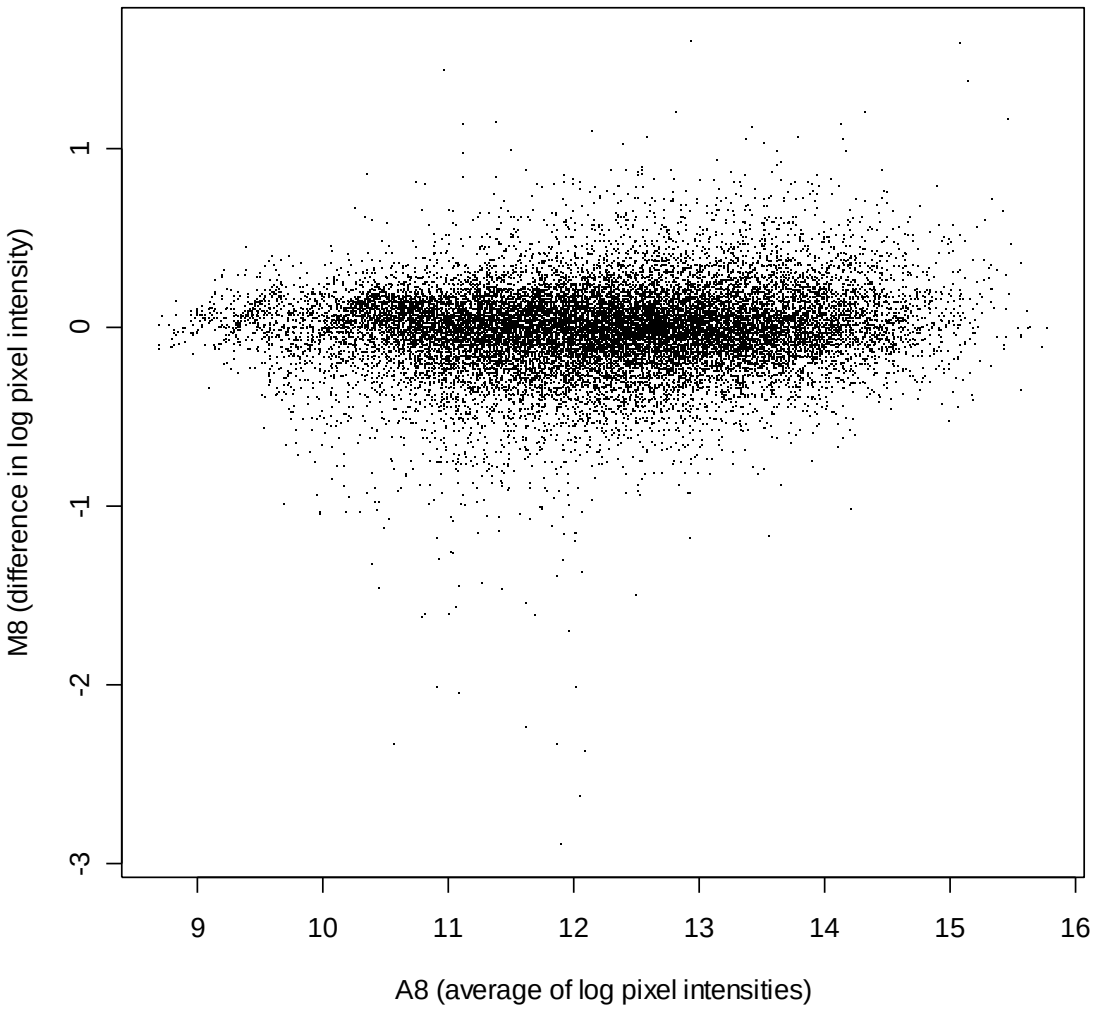

Figure 8. MvA Scatterplot for Microarray 8.
Where M = LogR-LogG, and A = ((LogR+LogG)/2). Log R equals the logarithmic value of the red intensity and Log G equals the logarithmic value of the green intensity. Pixel intensity equals the amount of signal from gene hybridization.

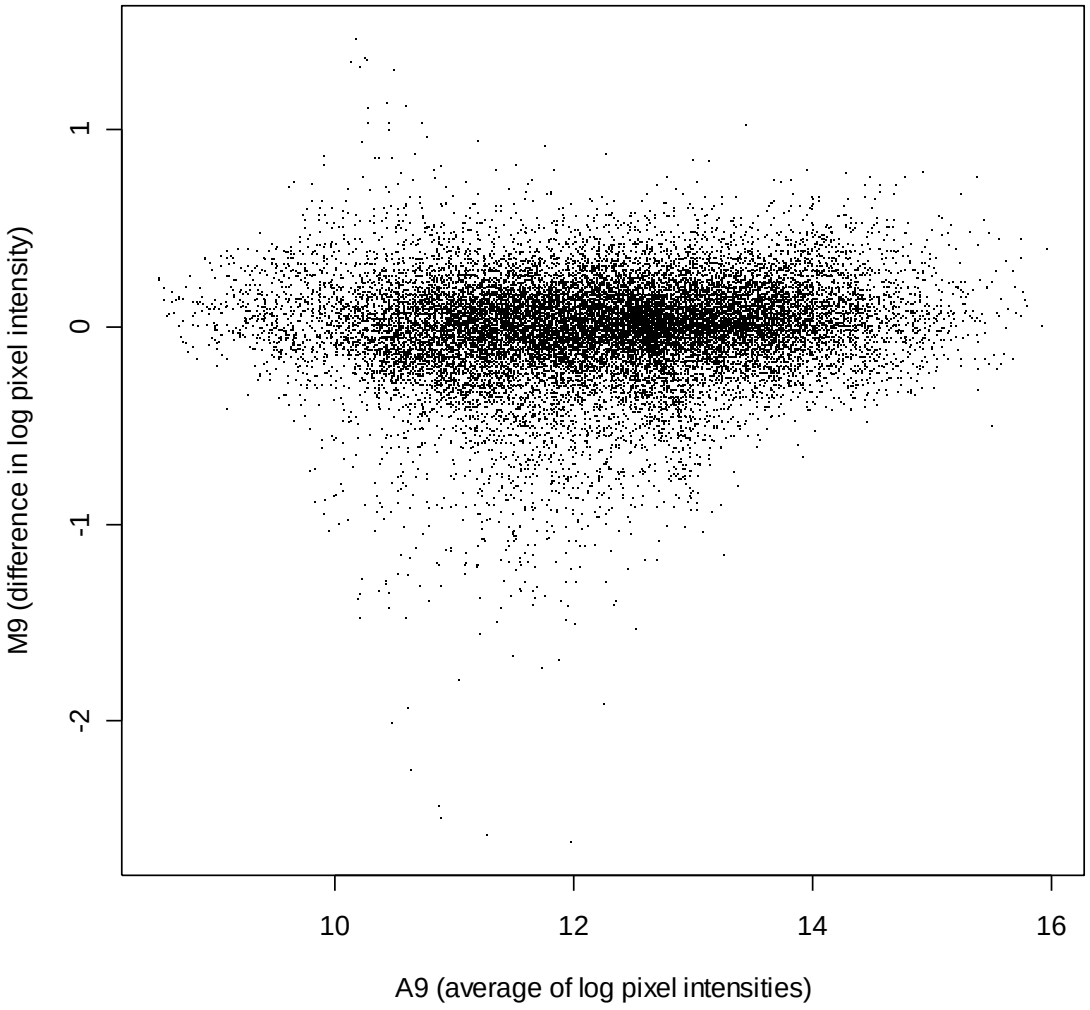

Figure 9. MvA Scatterplot for Microarray 9.
Where M = LogR-LogG, and A = ((LogR+LogG)/2). Log R equals the logarithmic value of the red intensity and Log G equals the logarithmic value of the green intensity. Pixel intensity equals the amount of signal from gene hybridization.

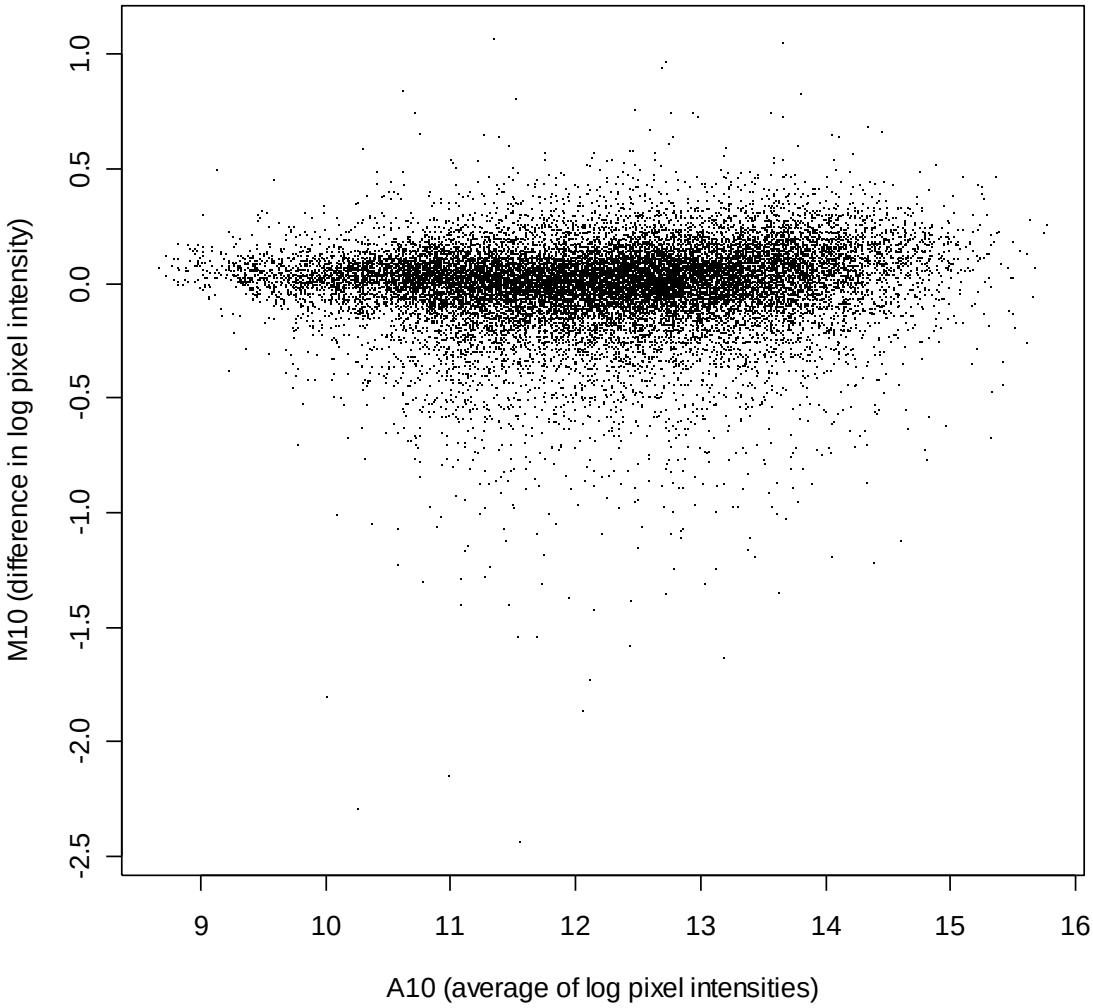

Figure 10. MvA Scatterplot for Microarray 10.
Where M = LogR-LogG, and A = ((LogR+LogG)/2). Log R equals the logarithmic value of the red intensity and Log G equals the logarithmic value of the green intensity. Pixel intensity equals the amount of signal from gene hybridization.

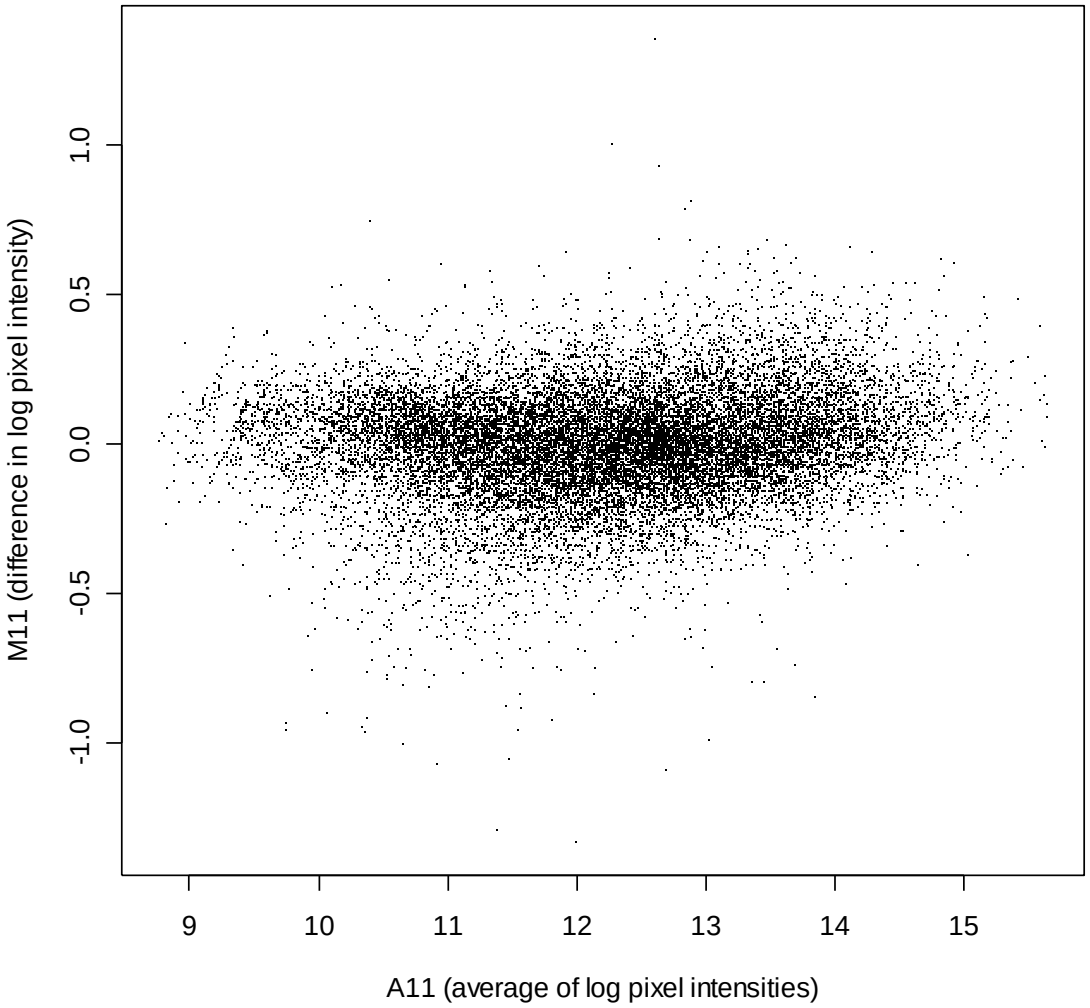

Figure 11. MvA Scatterplot for Microarray 11.
Where M = LogR-LogG, and A = ((LogR+LogG)/2). Log R equals the logarithmic value of the red intensity and Log G equals the logarithmic value of the green intensity. Pixel intensity equals the amount of signal from gene hybridization.

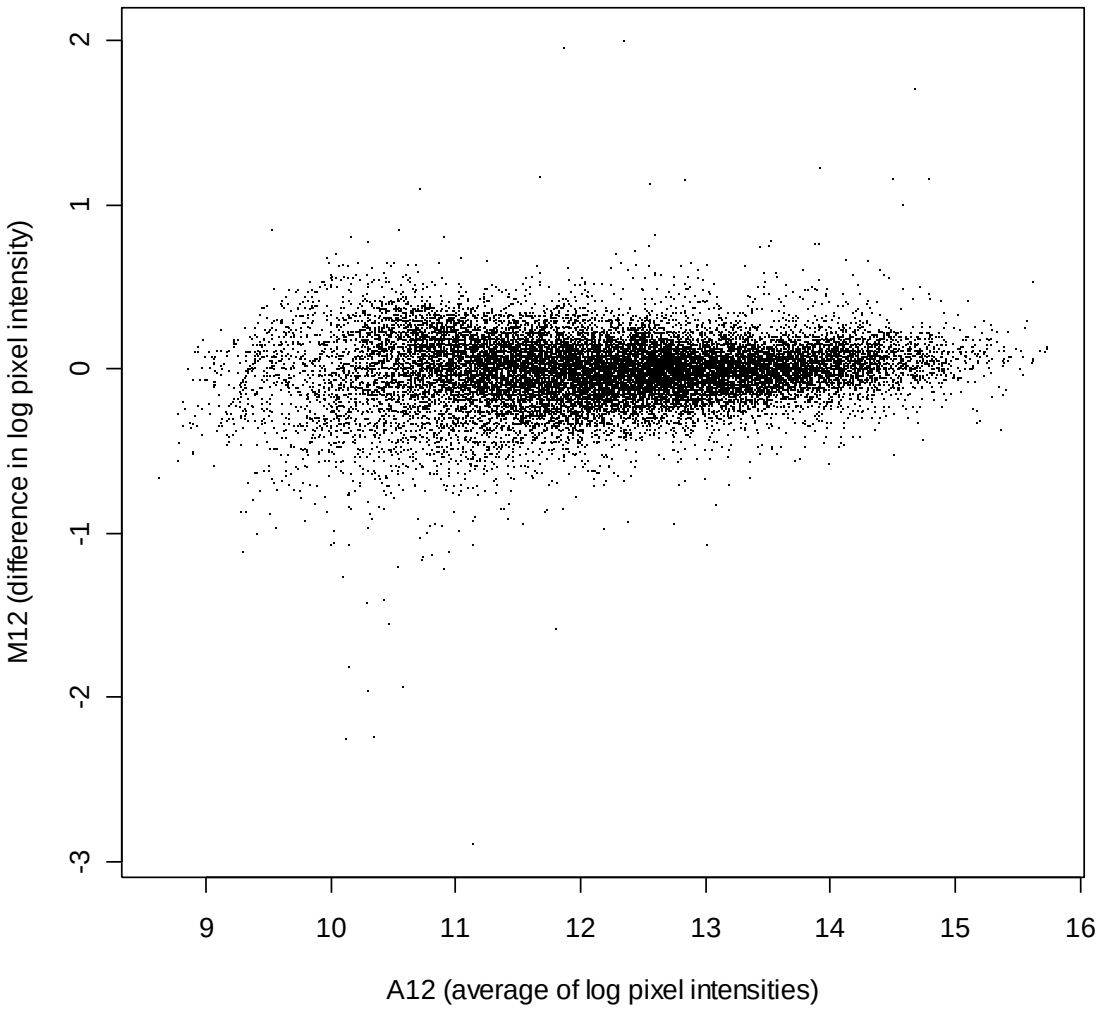

Figure 12. MvA Scatterplot for Microarray 12.
Where M = LogR-LogG, and A = ((LogR+LogG)/2). Log R equals the logarithmic value of the red intensity and Log G equals the logarithmic value of the green intensity. Pixel intensity equals the amount of signal from gene hybridization.

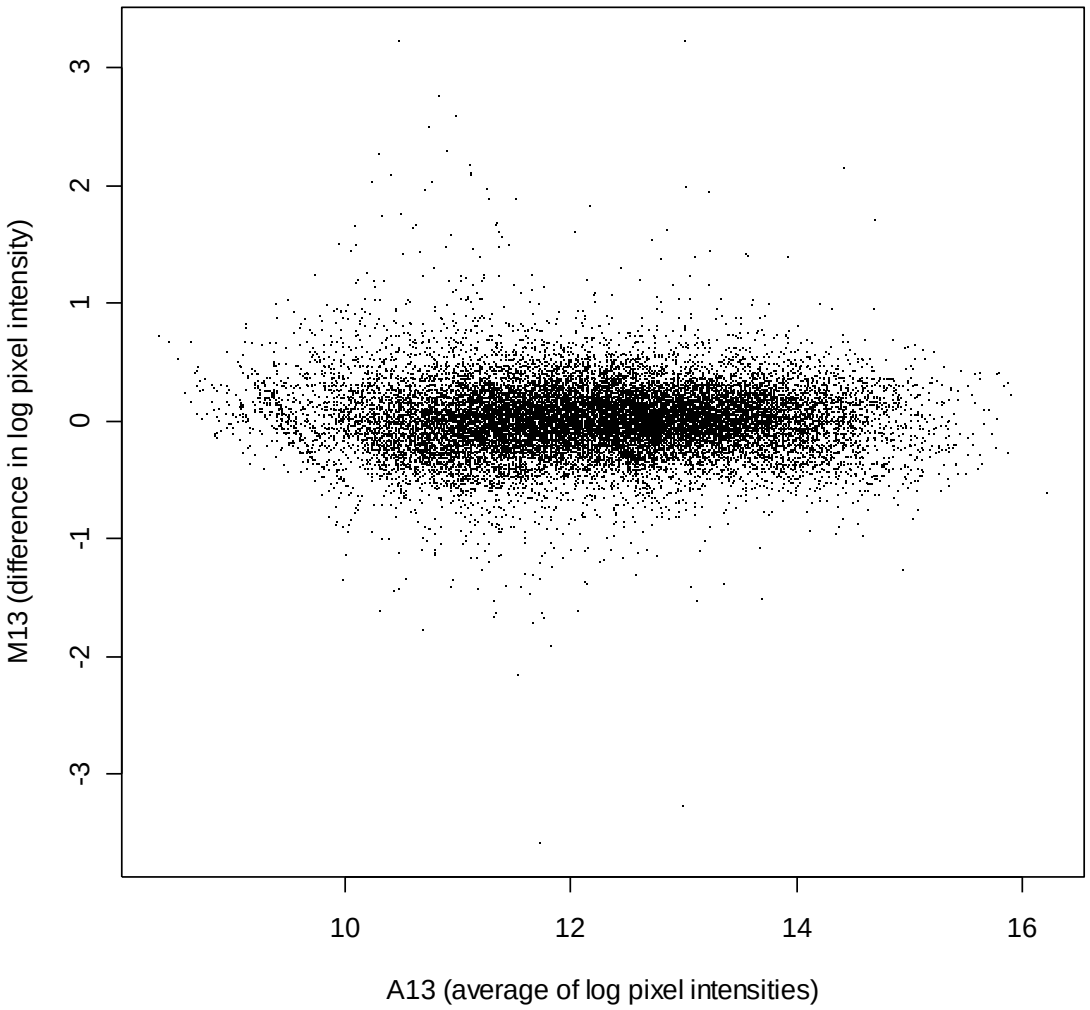

Figure 13. MvA Scatterplot for Microarray 13.
Where M = LogR-LogG, and A = ((LogR+LogG)/2). Log R equals the logarithmic value of the red intensity and Log G equals the logarithmic value of the green intensity. Pixel intensity equals the amount of signal from gene hybridization.

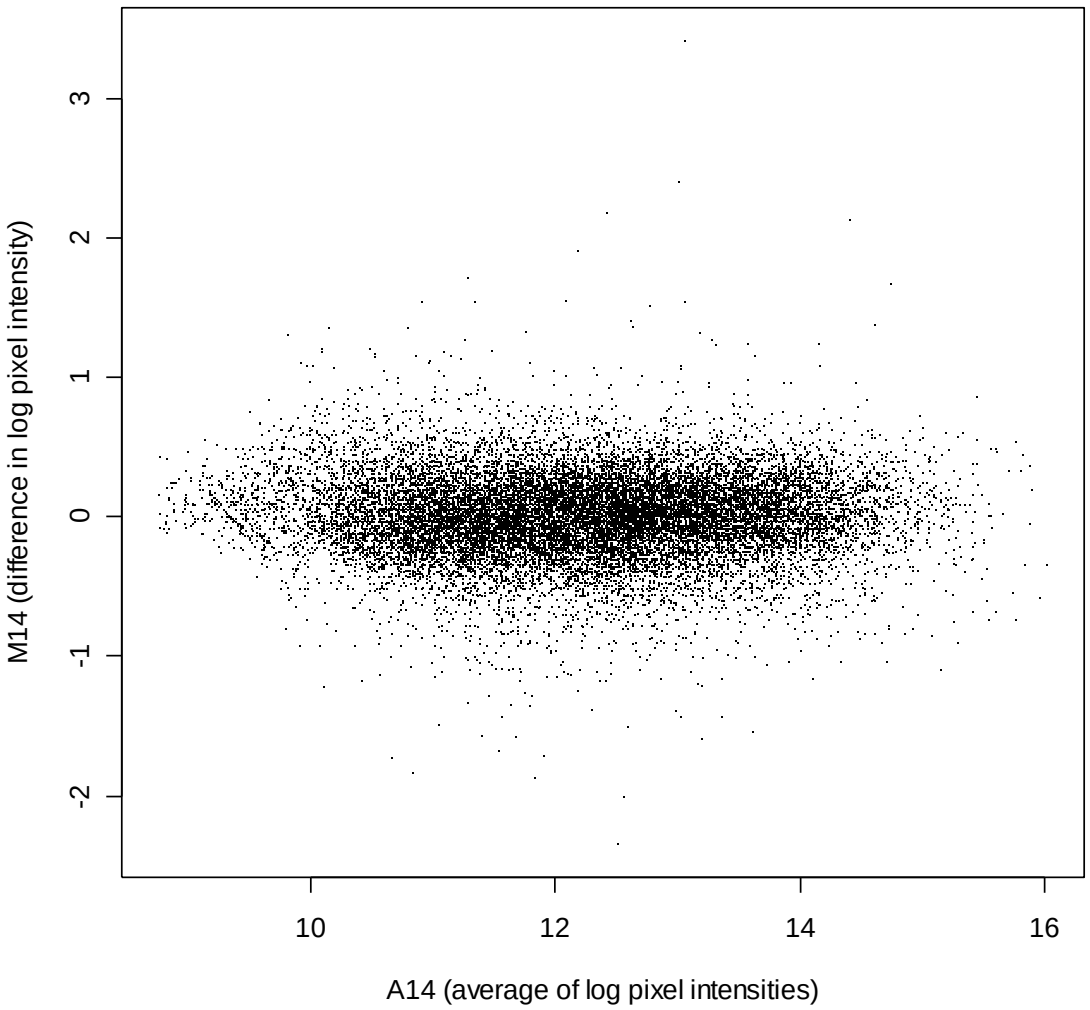

Figure 14. MvA Scatterplot for Microarray 14.
Where M = LogR-LogG, and A = ((LogR+LogG)/2). Log R equals the logarithmic value of the red intensity and Log G equals the logarithmic value of the green intensity. Pixel intensity equals the amount of signal from gene hybridization.

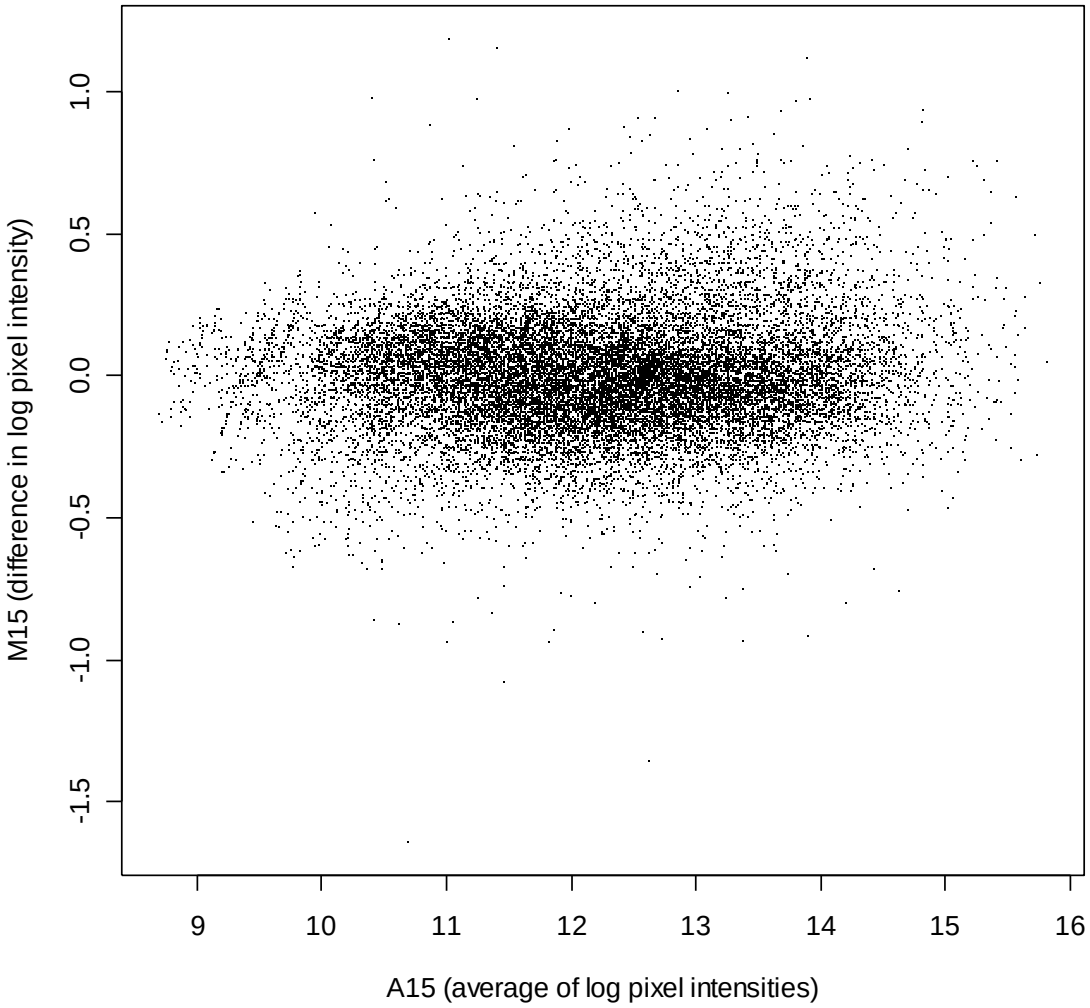

Figure 15. MvA Scatterplot for Microarray 15.
Where M = LogR-LogG, and A = ((LogR+LogG)/2). Log R equals the logarithmic value of the red intensity and Log G equals the logarithmic value of the green intensity. Pixel intensity equals the amount of signal from gene hybridization.

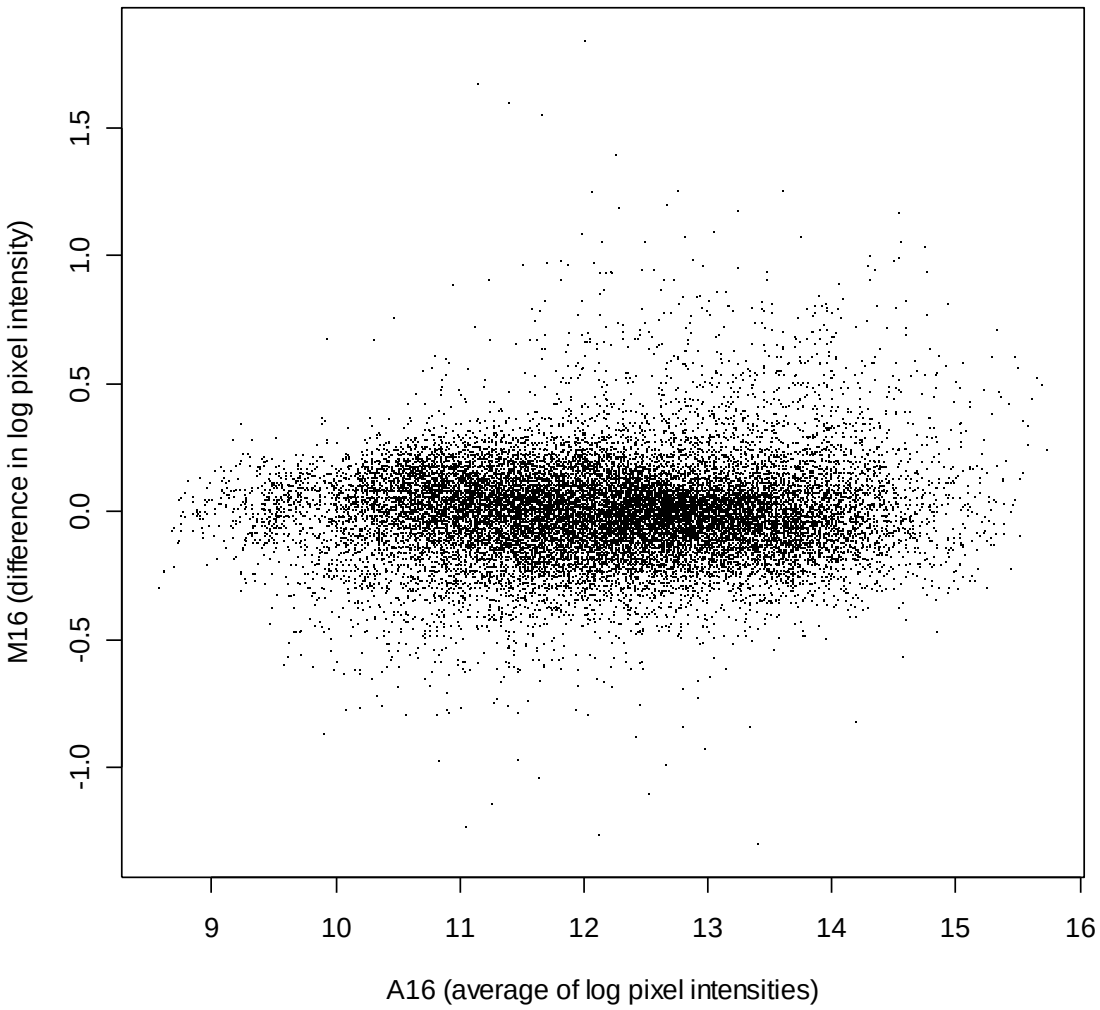

Figure 16. MvA Scatterplot for Microarray 16.
Where M = LogR-LogG, and A = ((LogR+LogG)/2). Log R equals the logarithmic value of the red intensity and Log G equals the logarithmic value of the green intensity. Pixel intensity equals the amount of signal from gene hybridization.

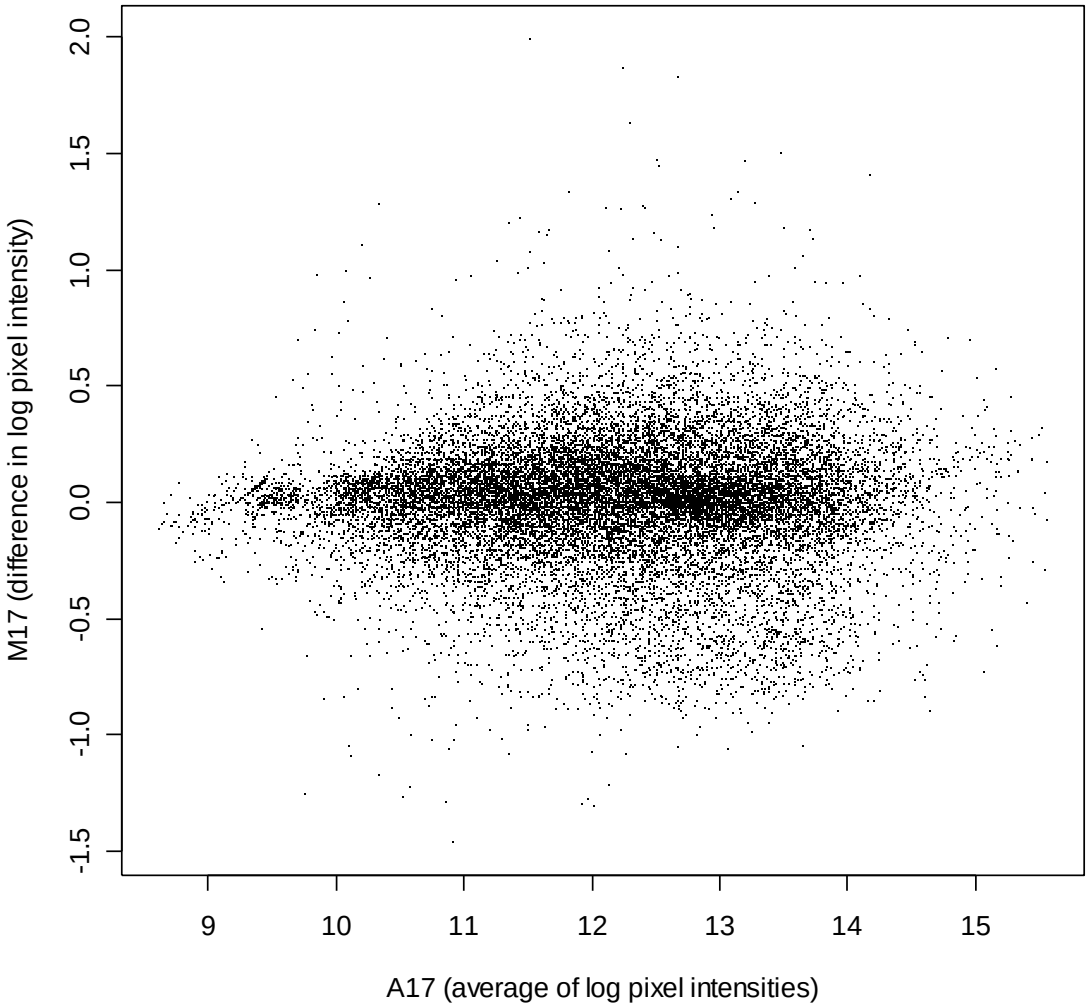

Figure 17. MvA Scatterplot for Microarray 17.
Where M = LogR-LogG, and A = ((LogR+LogG)/2). Log R equals the logarithmic value of the red intensity and Log G equals the logarithmic value of the green intensity. Pixel intensity equals the amount of signal from gene hybridization.

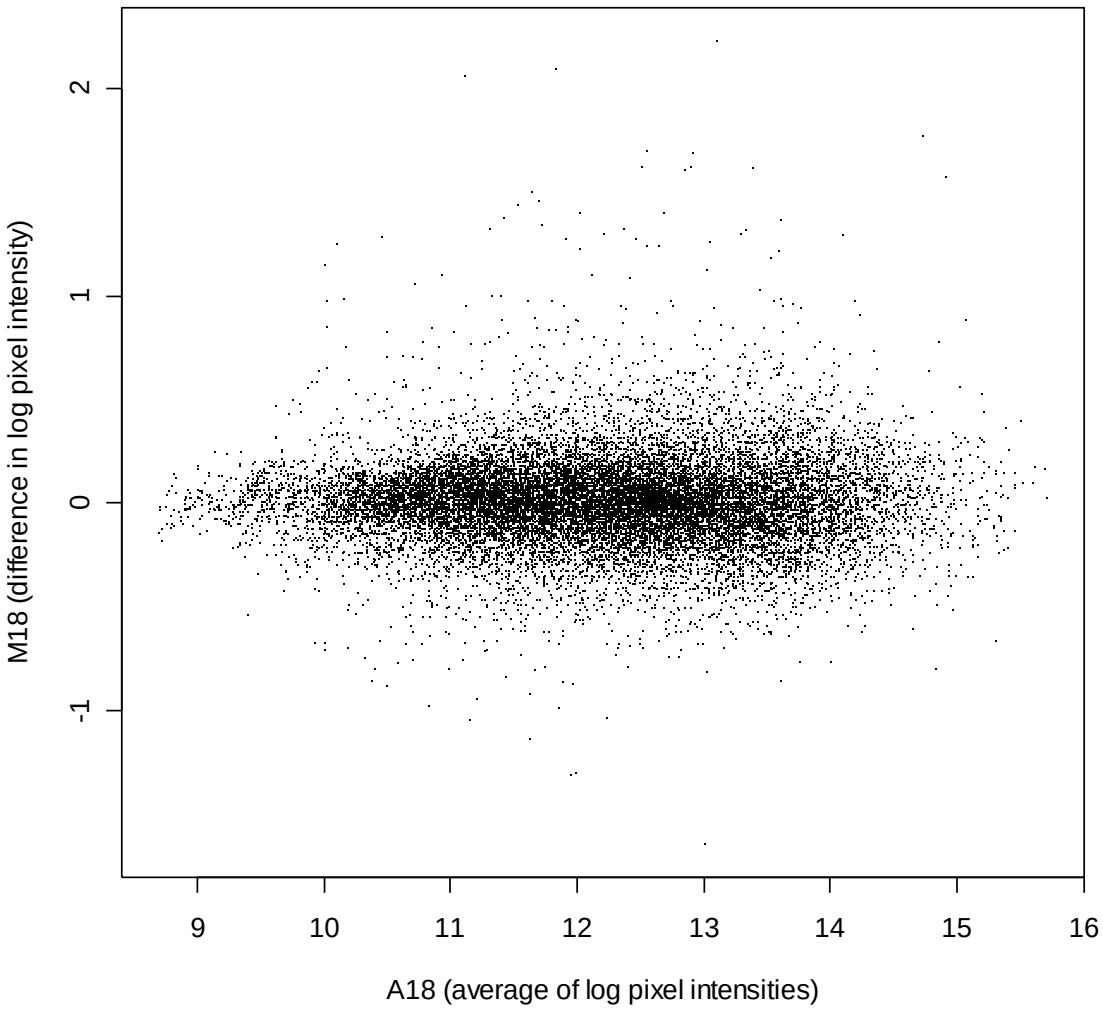

Figure 18. MvA Scatterplot for Microarray 18.
Where M = LogR-LogG, and A = ((LogR+LogG)/2). Log R equals the logarithmic value of the red intensity and Log G equals the logarithmic value of the green intensity. Pixel intensity equals the amount of signal from gene hybridization.

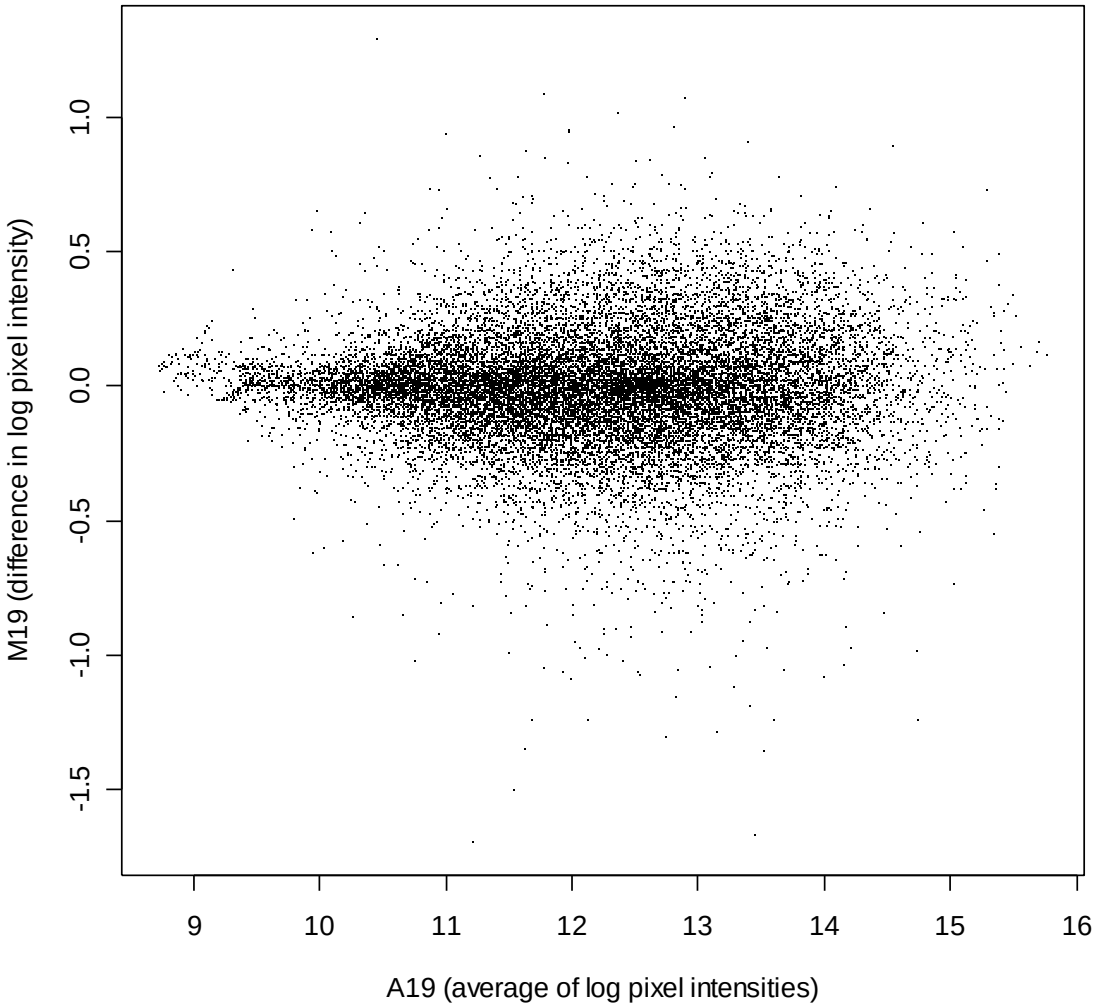

Figure 19. MvA Scatterplot for Microarray 19.
Where M = LogR-LogG, and A = ((LogR+LogG)/2). Log R equals the logarithmic value of the red intensity and Log G equals the logarithmic value of the green intensity. Pixel intensity equals the amount of signal from gene hybridization.

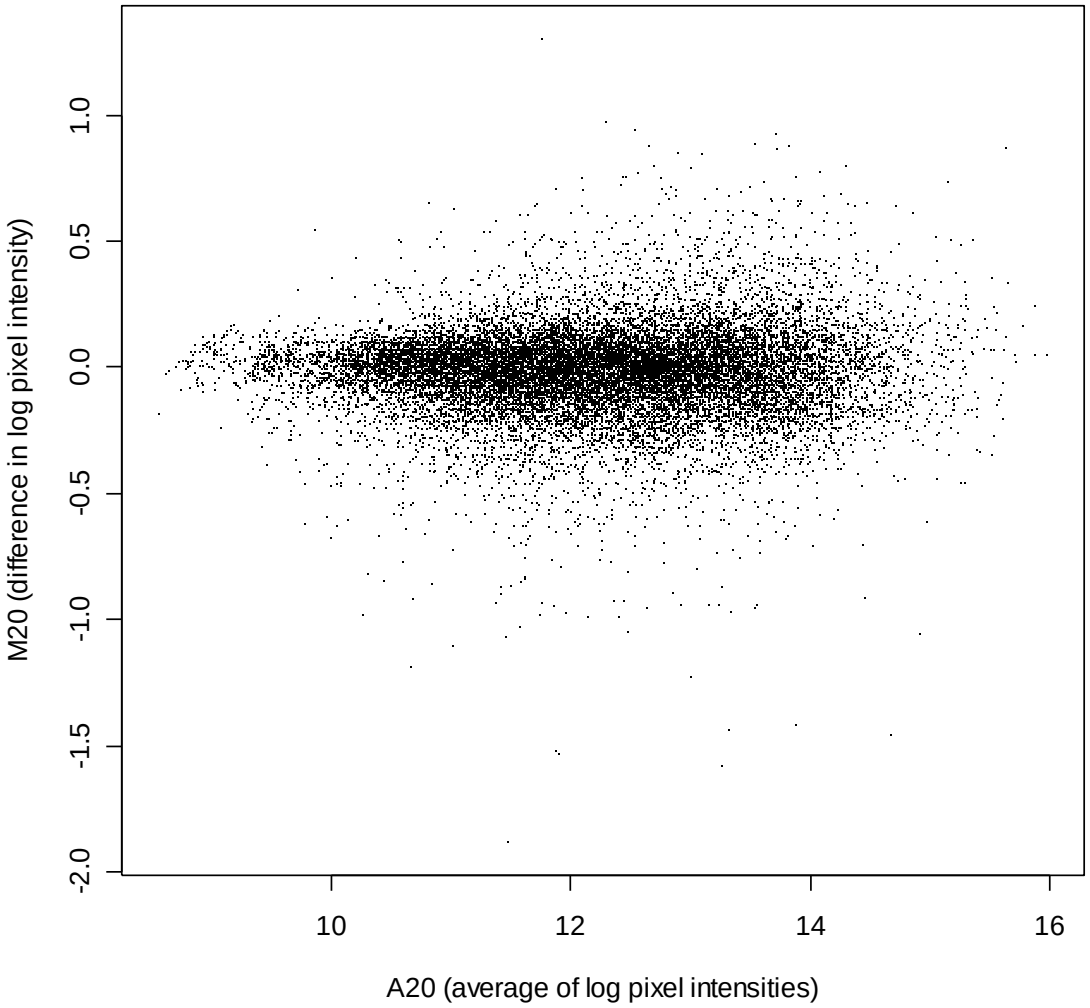

Figure 20. MvA Scatterplot for Microarray 20.
Where M = LogR-LogG, and A = ((LogR+LogG)/2). Log R equals the logarithmic value of the red intensity and Log G equals the logarithmic value of the green intensity. Pixel intensity equals the amount of signal from gene hybridization.

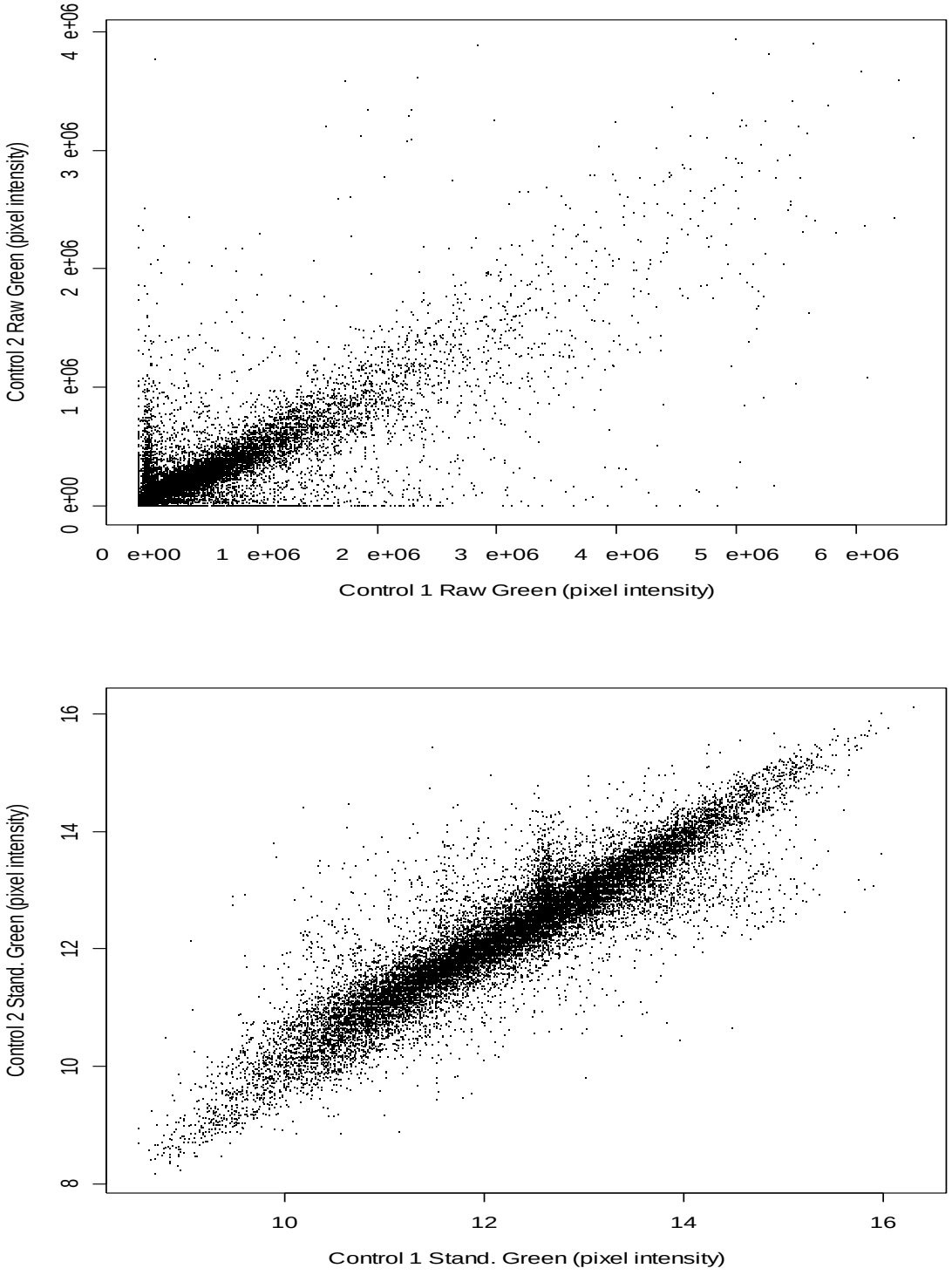

Figure 21. Scatterplot of Green Intensities for Dye Duplicates in Microarrays 1 and 2. For microarrays 1 and 2, both control samples were labeled with Cy3 (green) fluorescent dye. In this scatterplot, both pixel intensities were compared to ensure uniform and effective labeling. Pixel intensity equals the amount of signal from gene hybridization.

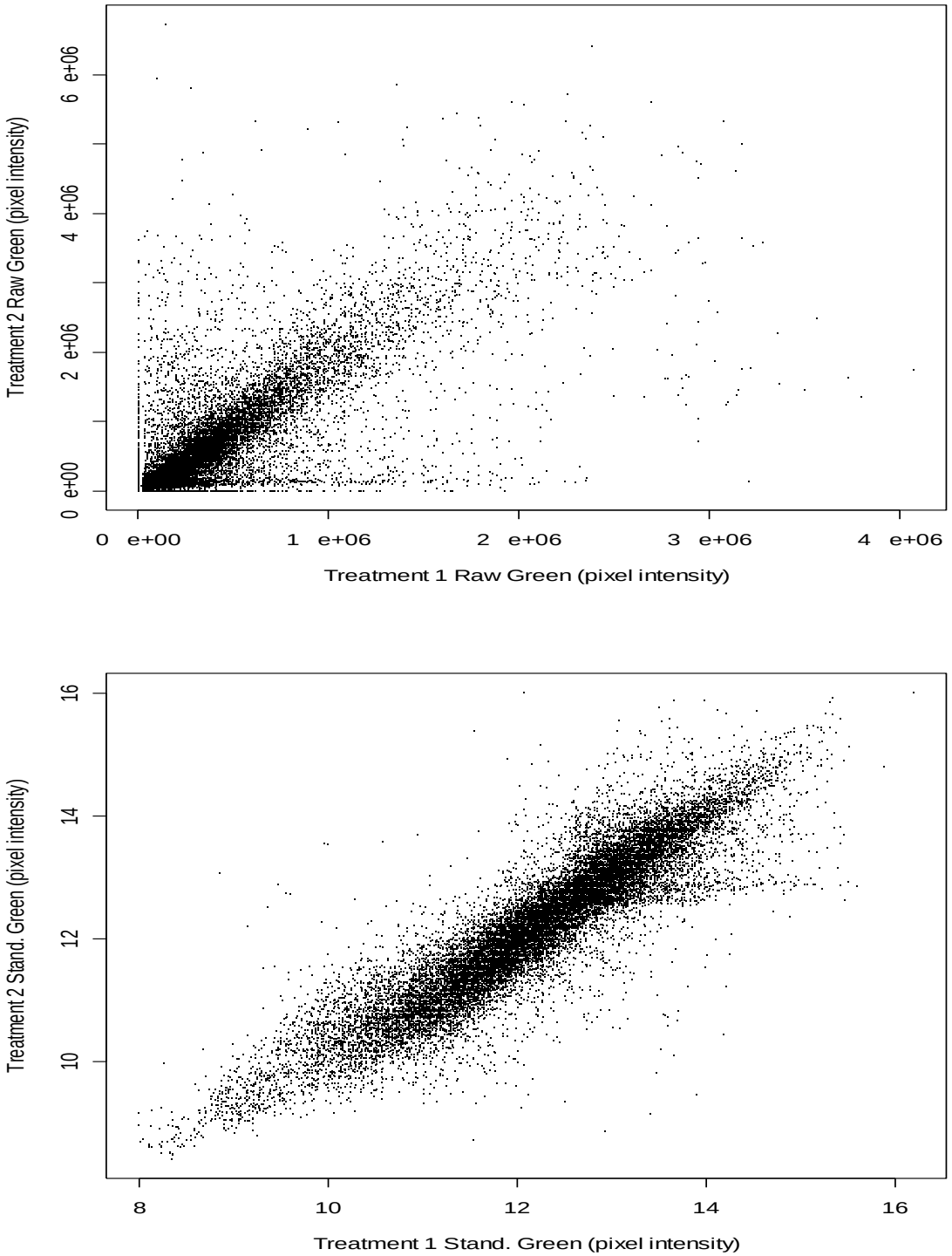

Figure 22. Scatterplot of Green Intensities for Dye Duplicates in Microarrays 3 and 4. For microarrays 3 and 4, both treatment samples were labeled with Cy3 (green) fluorescent dye. In this scatterplot, both pixel intensities were compared to ensure uniform and effective labeling. Pixel intensity equals the amount of signal from gene hybridization.

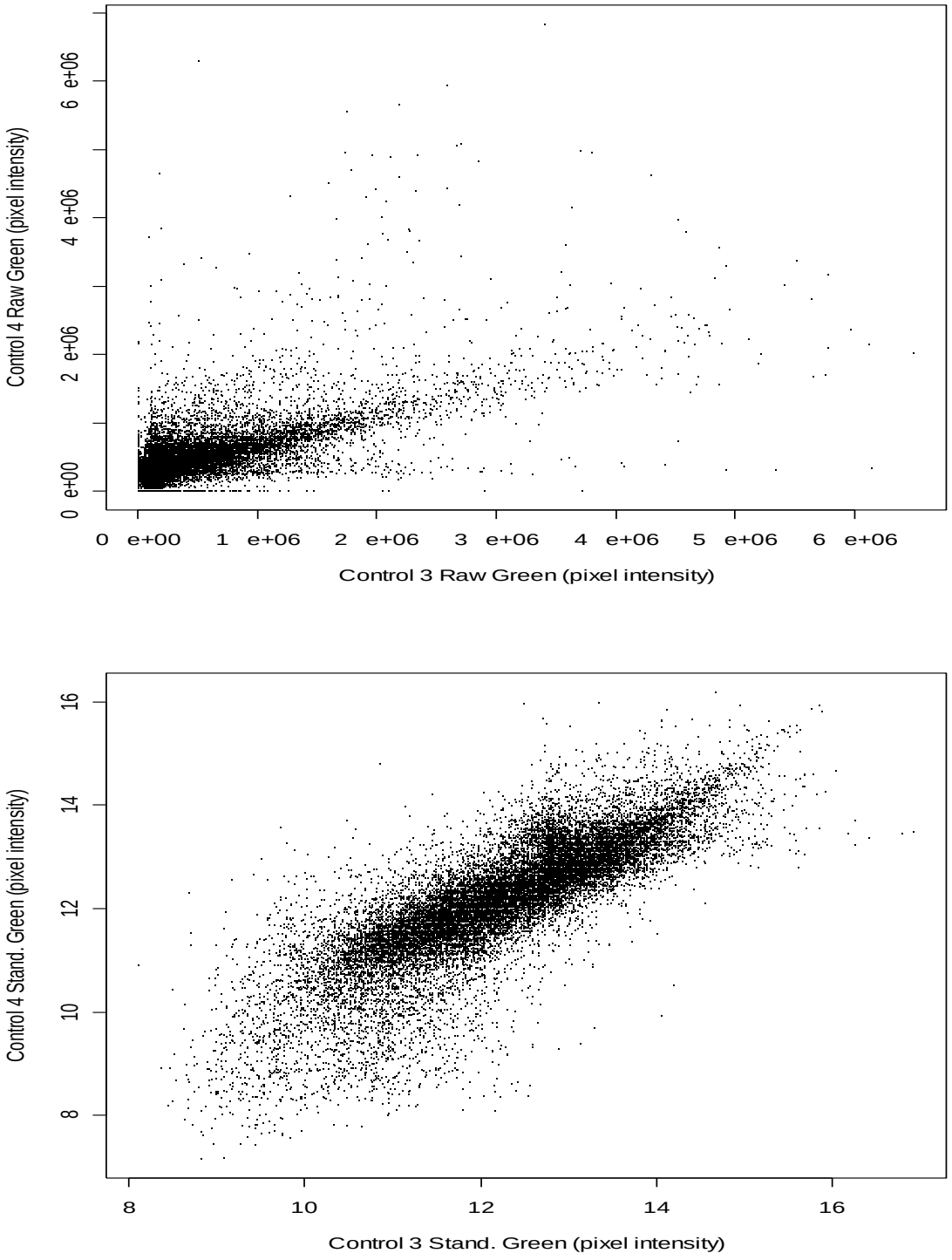

Figure 23. Scatterplot of Green Intensities for Dye Duplicates in Microarrays 5 and 6. For microarrays 5 and 6, both control samples were labeled with Cy3 (green) fluorescent dye. In this scatterplot, both pixel intensities were compared to ensure uniform and effective labeling. Pixel intensity equals the amount of signal from gene hybridization.

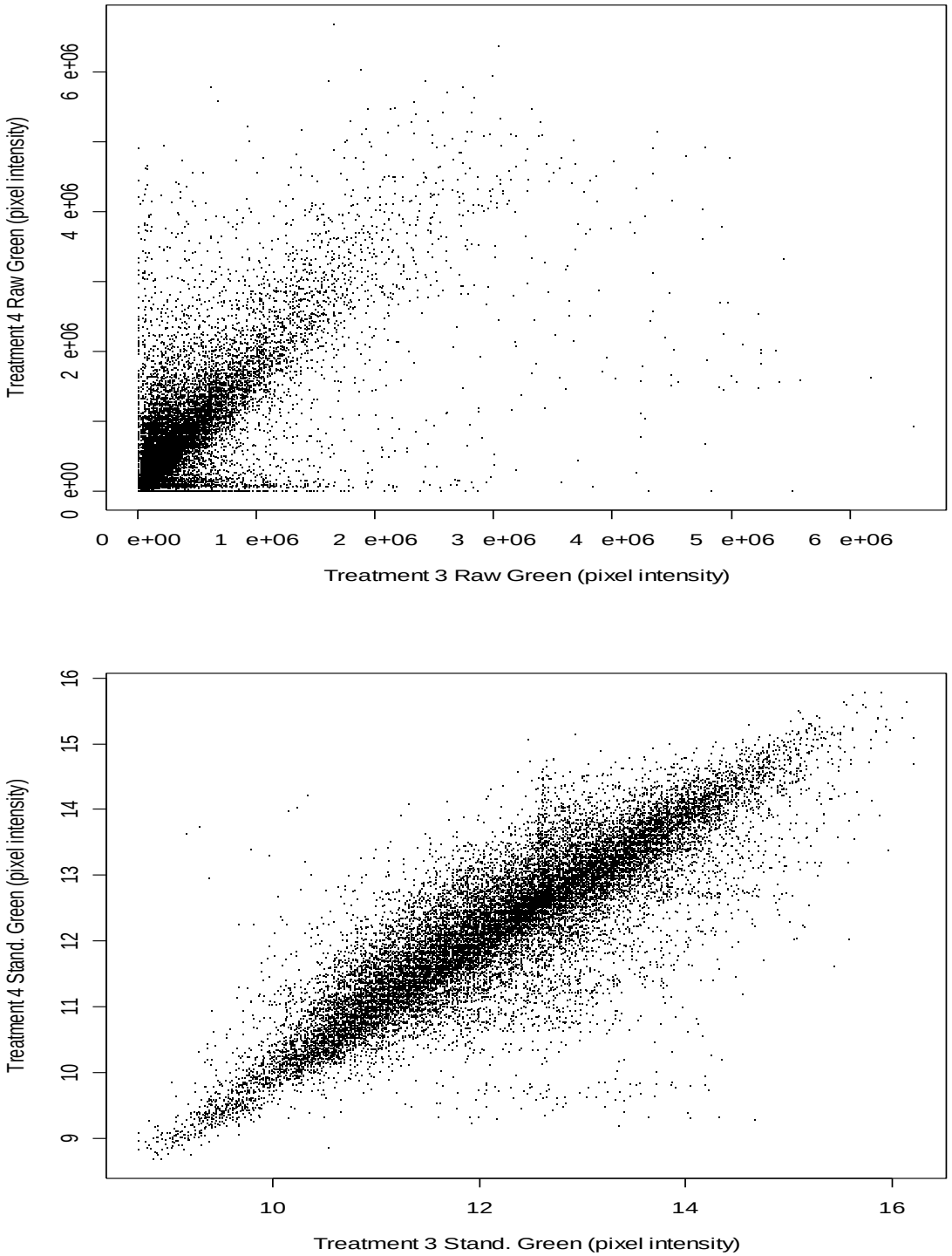

Figure 24. Scatterplot of Green Intensities for Dye Duplicates in Microarrays 7 and 8. For microarrays 7 and 8, both treatment samples were labeled with Cy3 (green) fluorescent dye. In this scatterplot, both pixel intensities were compared to ensure uniform and effective labeling. Pixel intensity equals the amount of signal from gene hybridization.

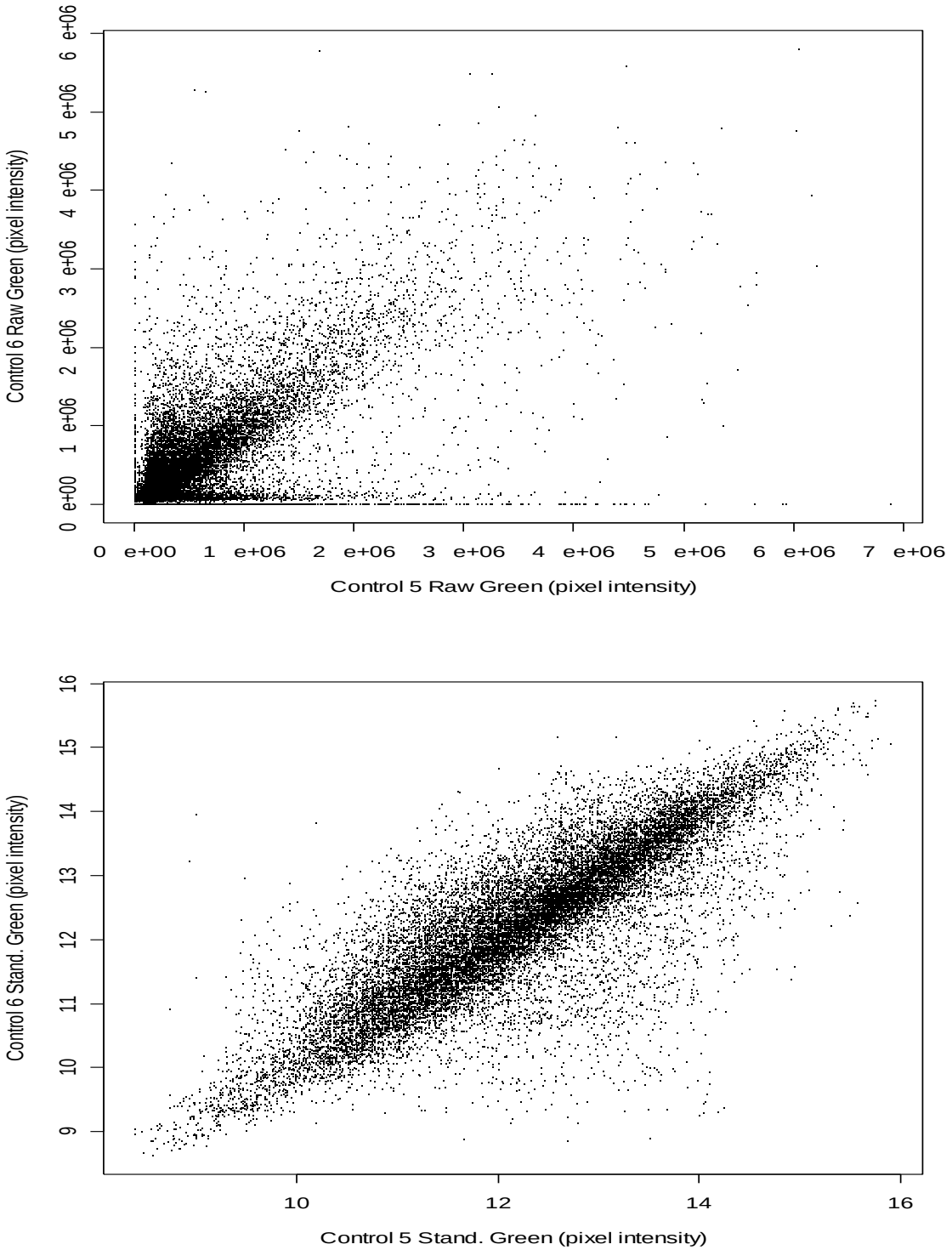

Figure 25. Scatterplot of Green Intensities for Dye Duplicates in Microarrays 9 and 10. For microarrays 9 and 10, both control samples were labeled with Cy3 (green) fluorescent dye. In this scatterplot, both pixel intensities were compared to ensure uniform and effective labeling. Pixel intensity equals the amount of signal from gene hybridization.

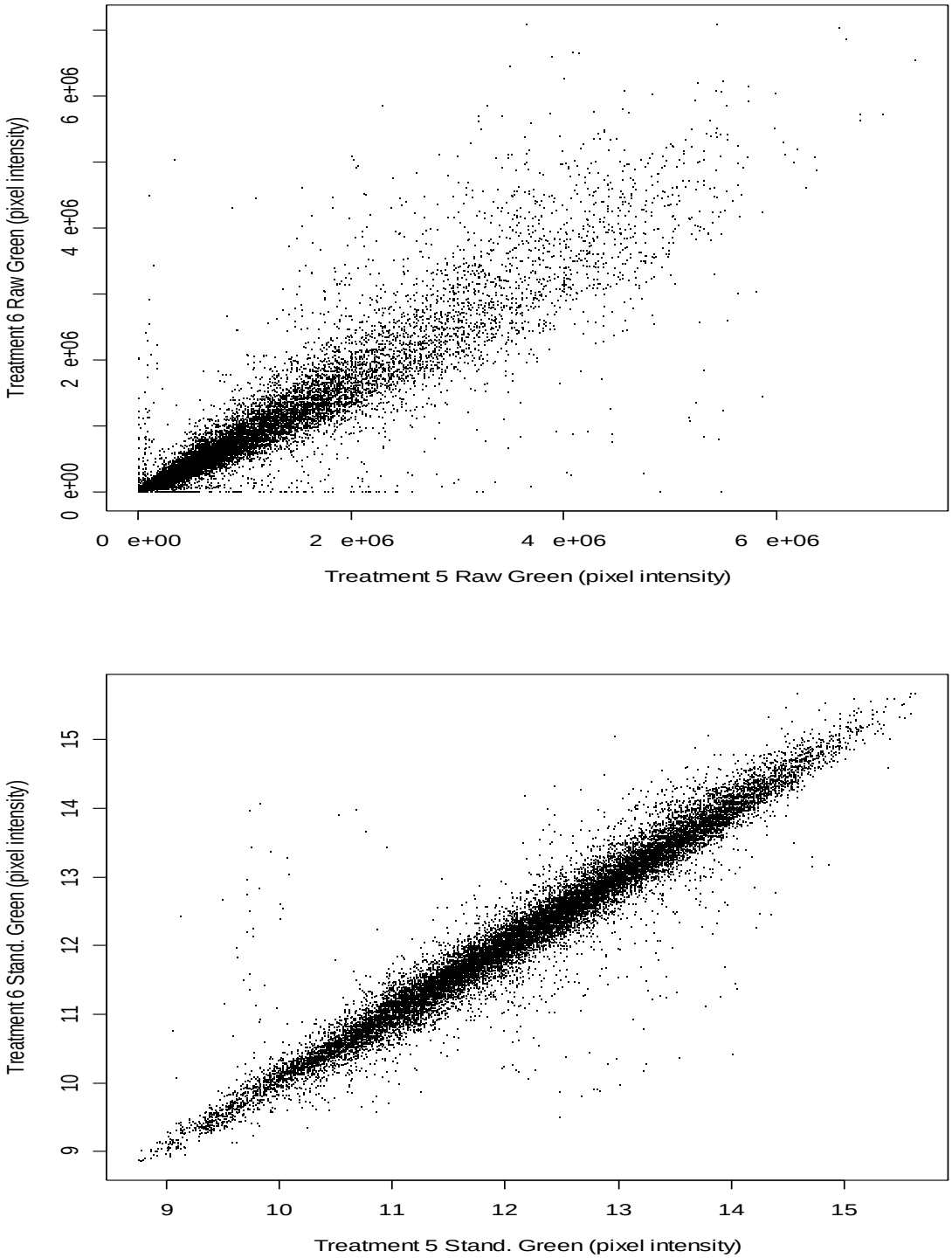

Figure 26. Scatterplot of Green Intensities for Dye Duplicates in Microarrays 11 and 12. For microarrays 11 and 12, both treatment samples were labeled with Cy3 (green) fluorescent dye. In this scatterplot, both pixel intensities were compared to ensure uniform and effective labeling. Pixel intensity equals the amount of signal from gene hybridization.

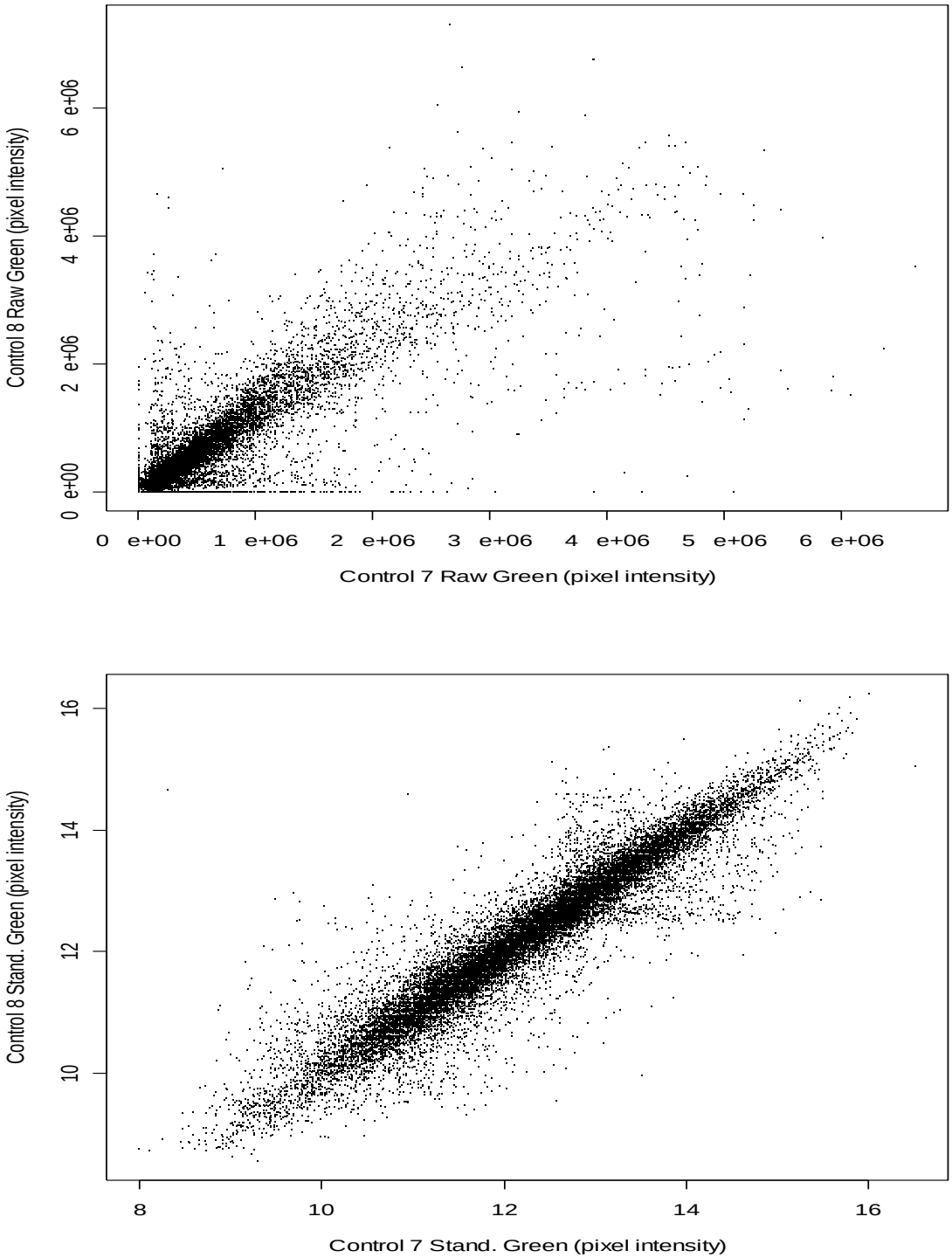

Figure 27. Scatterplot of Green Intensities for Dye Duplicates in Microarrays 13 and 14. For microarrays 13 and 14, both control samples were labeled with Cy3 (green) fluorescent dye. In this scatterplot, both pixel intensities were compared to ensure uniform and effective labeling. Pixel intensity equals the amount of signal from gene hybridization.

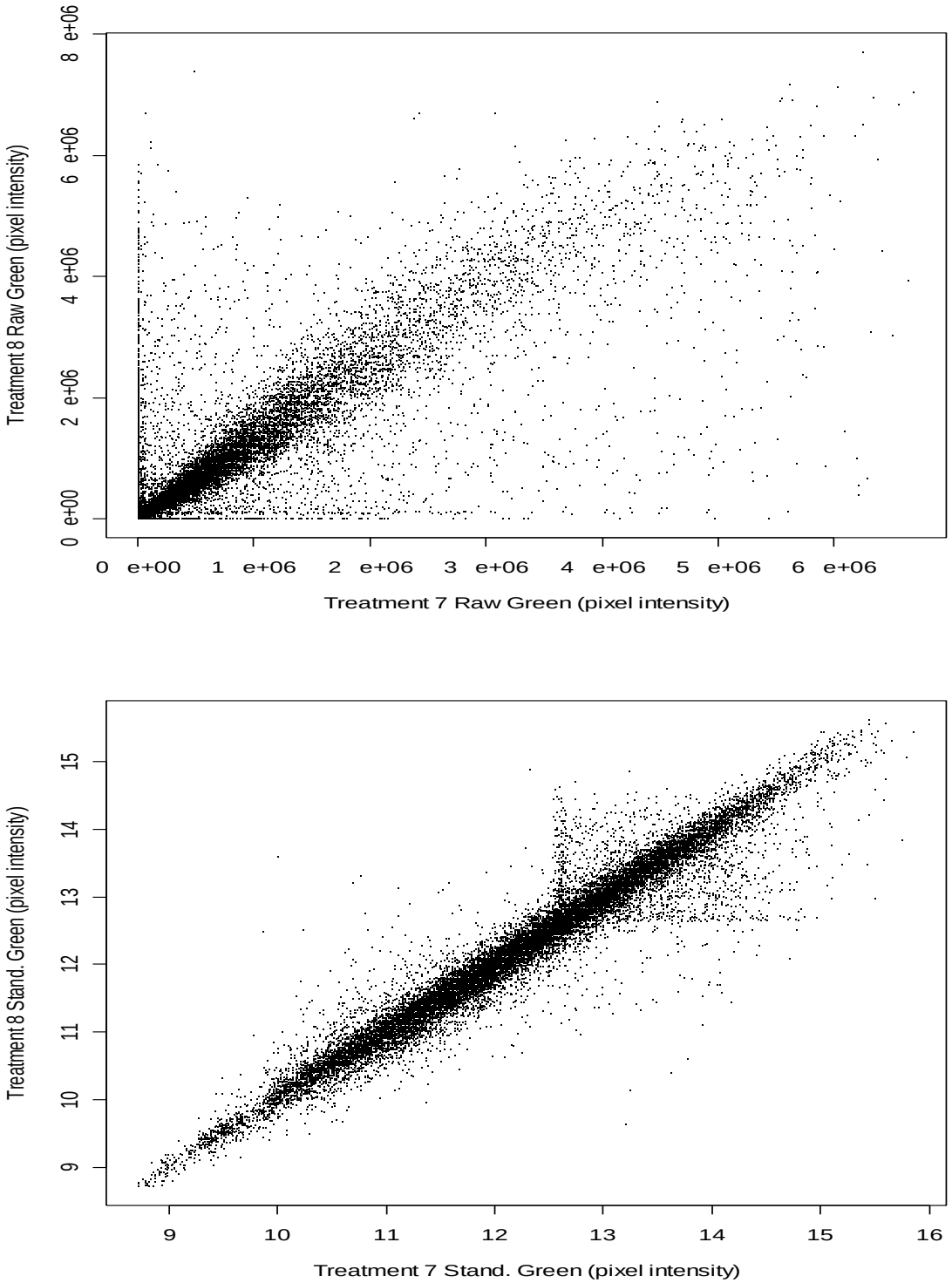

Figure 28. Scatterplot of Green Intensities for Dye Duplicates in Microarrays 15 and 16. For microarrays 15 and 16, both treatment samples were labeled with Cy3 (green) fluorescent dye. In this scatterplot, both pixel intensities were compared to ensure uniform and effective labeling. Pixel intensity equals the amount of signal from gene hybridization.

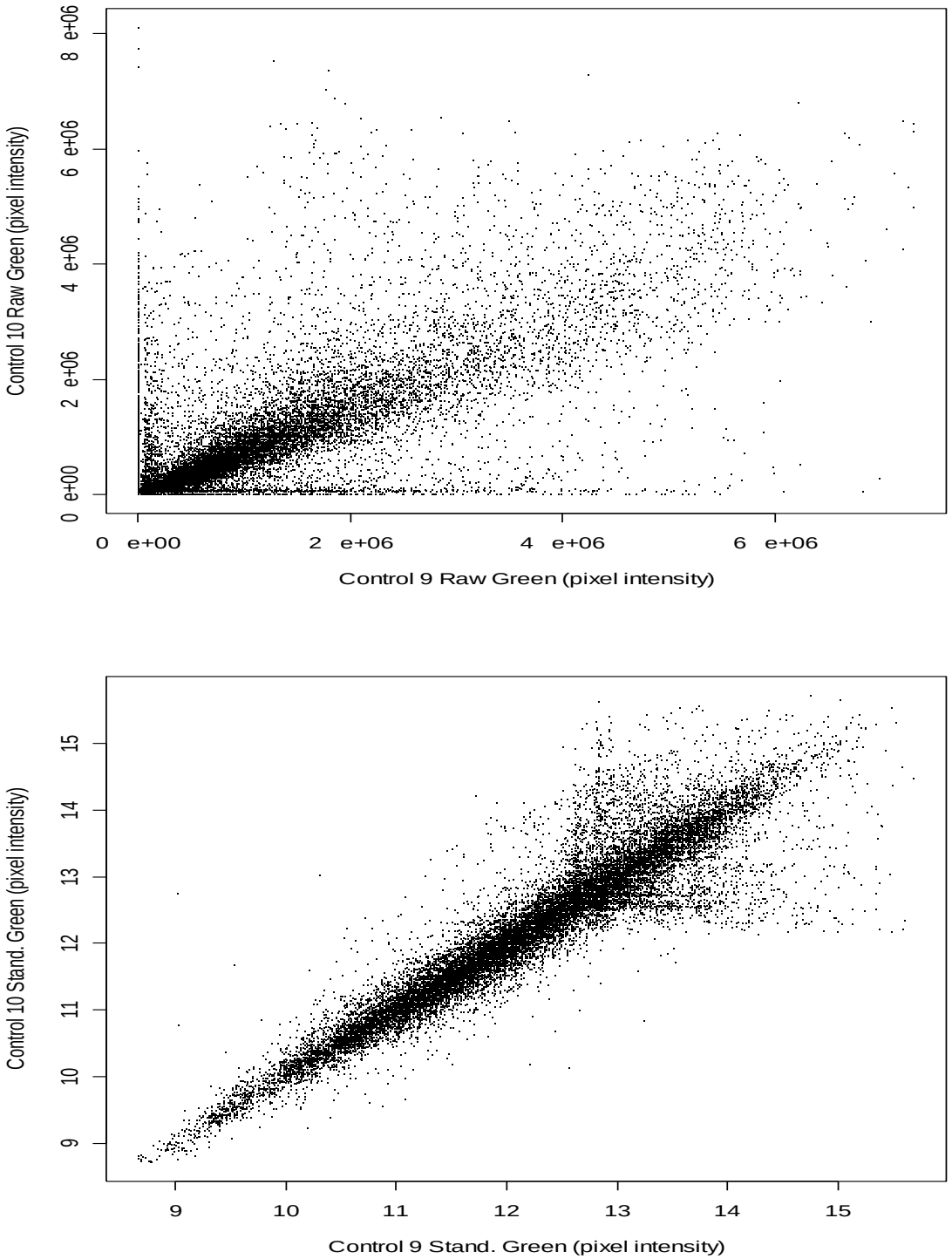

Figure 29. Scatterplot of Green Intensities for Dye Duplicates in Microarrays 17 and 18. For microarrays 17 and 18, both control samples were labeled with Cy3 (green) fluorescent dye. In this scatterplot, both pixel intensities were compared to ensure uniform and effective labeling. Pixel intensity equals the amount of signal from gene hybridization.

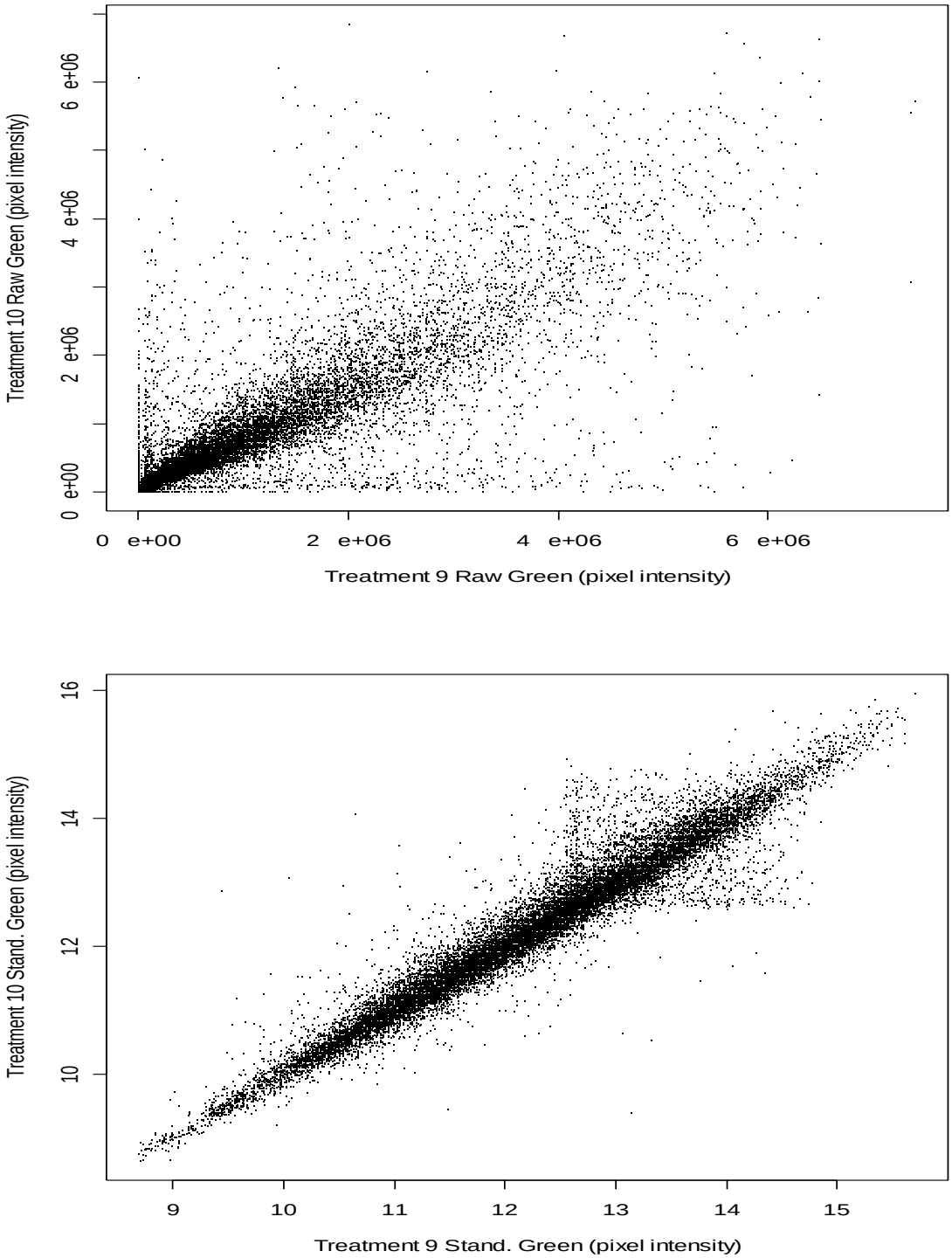

Figure 30. Scatterplot of Green Intensities for Dye Duplicates in Microarrays 19 and 20. For microarrays 19 and 20, both treatment samples were labeled with Cy3 (green) fluorescent dye. In this scatterplot, both pixel intensities were compared to ensure uniform and effective labeling. Pixel intensity equals the amount of signal from gene hybridization.

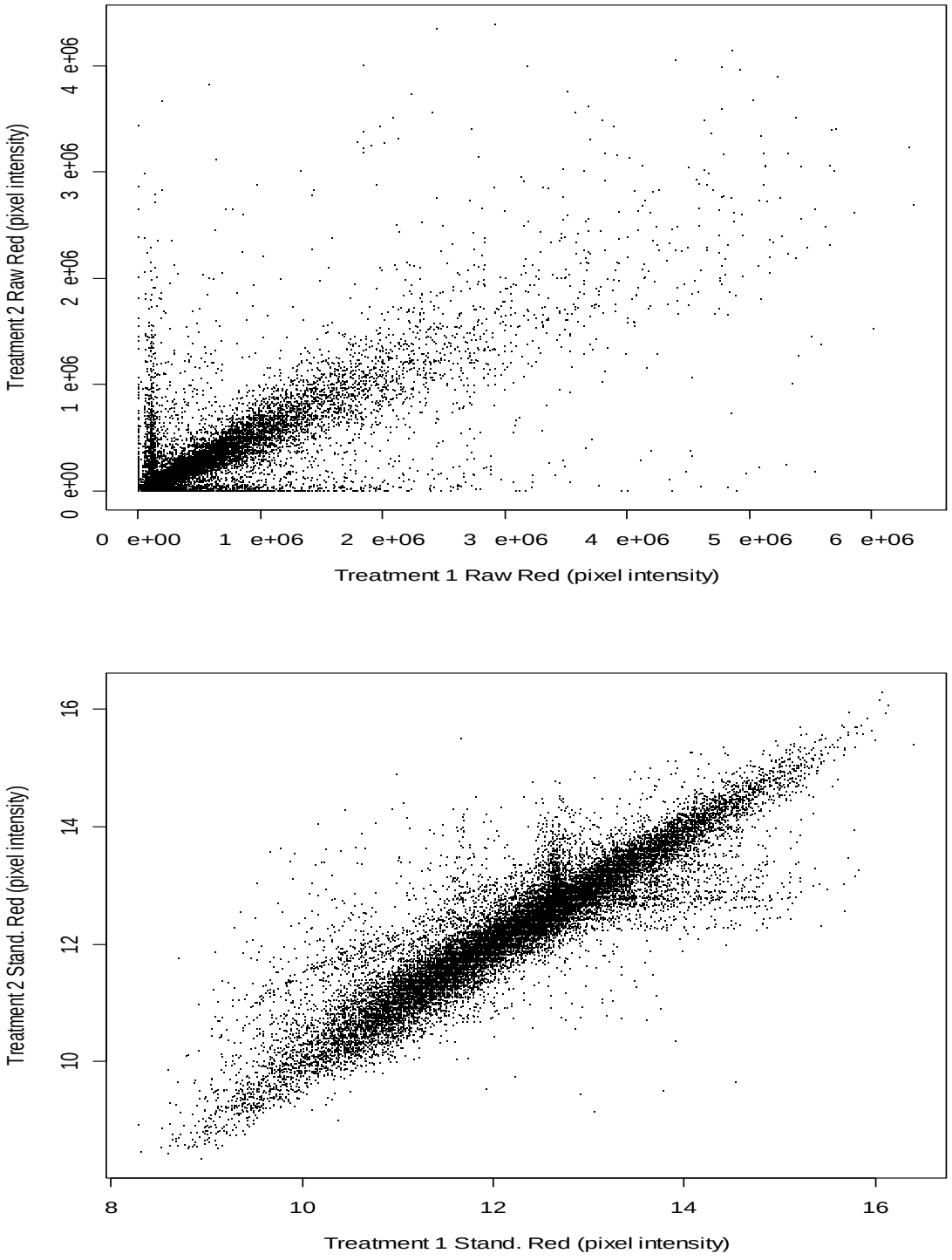

Figure 31. Scatterplot of Red Intensities for Dye Duplicates in Microarrays 1 and 2. For microarrays 1 and 2, both treatment samples were labeled with Cy5 (red) fluorescent dye. In this scatterplot, both pixel intensities were compared to ensure uniform and effective labeling. Pixel intensity equals the amount of signal from gene hybridization.

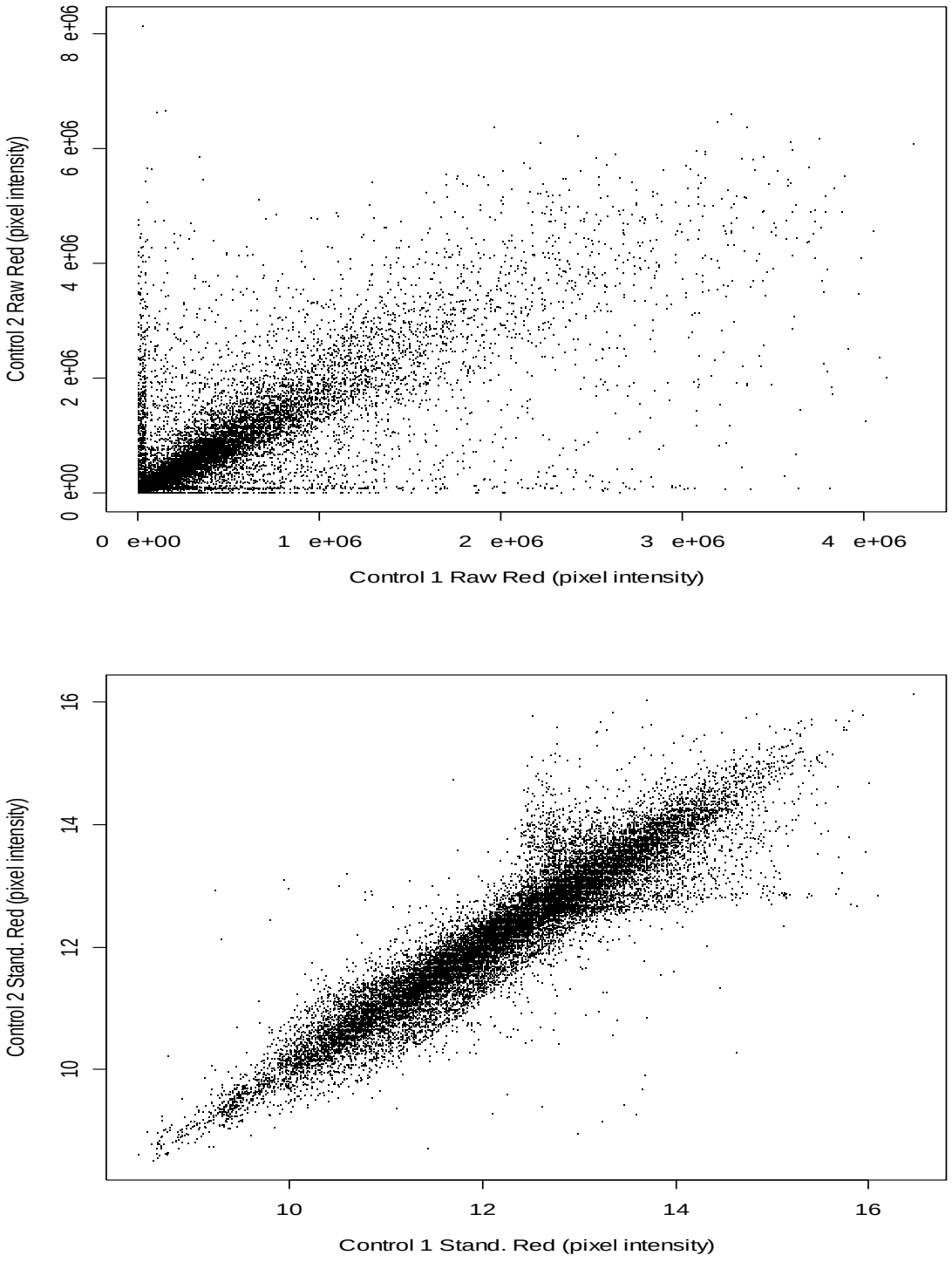

Figure 32. Scatterplot of Red Intensities for Dye Duplicates in Microarrays 3 and 4. For microarrays 3 and 4, both control samples were labeled with Cy5 (red) fluorescent dye. In this scatterplot, both pixel intensities were compared to ensure uniform and effective labeling. Pixel intensity equals the amount of signal from gene hybridization.

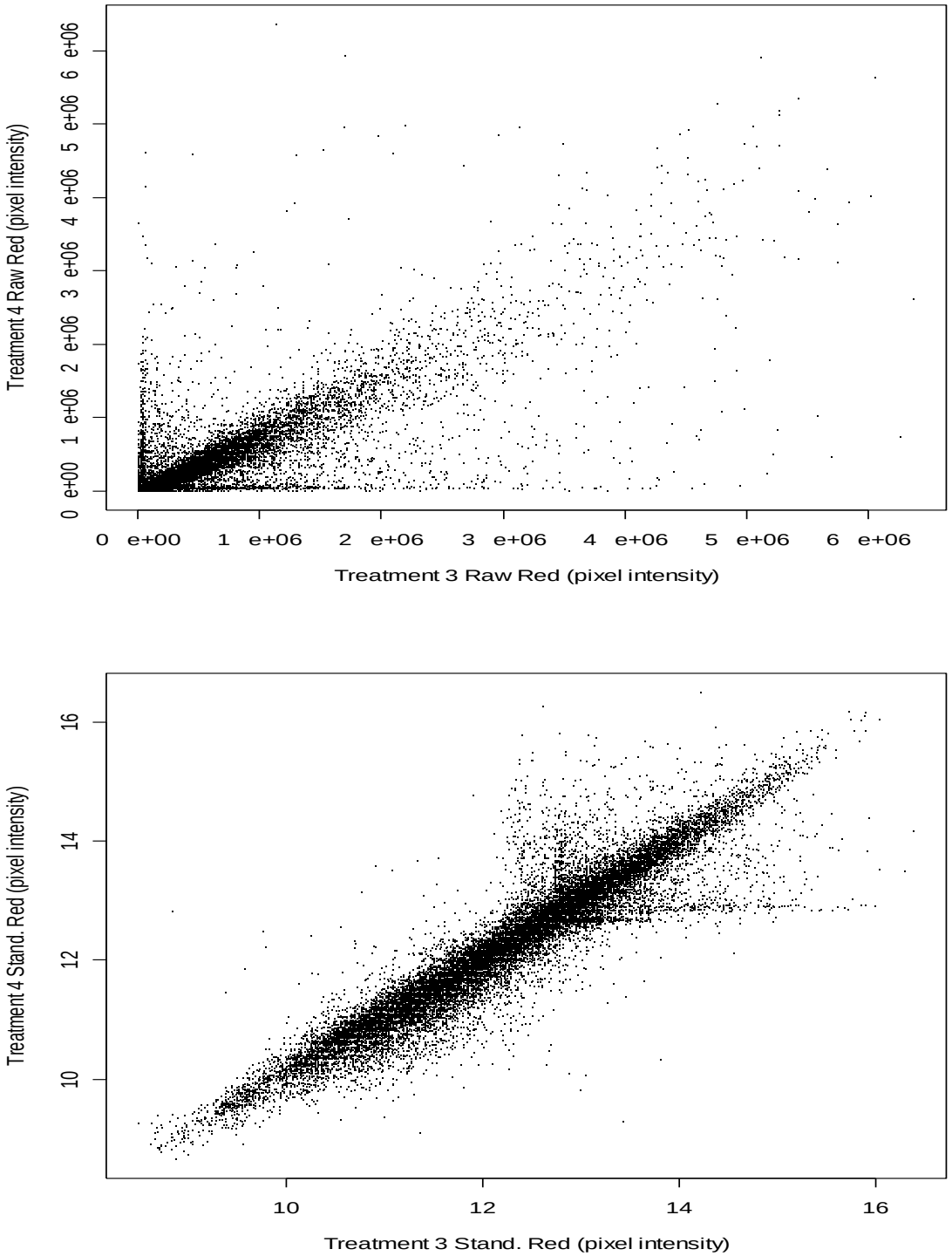

Figure 33. Scatterplot of Red Intensities for Dye Duplicates in Microarrays 5 and 6. For microarrays 5 and 6, both treatment samples were labeled with Cy5 (red) fluorescent dye. In this scatterplot, both pixel intensities were compared to ensure uniform and effective labeling. Pixel intensity equals the amount of signal from gene hybridization.

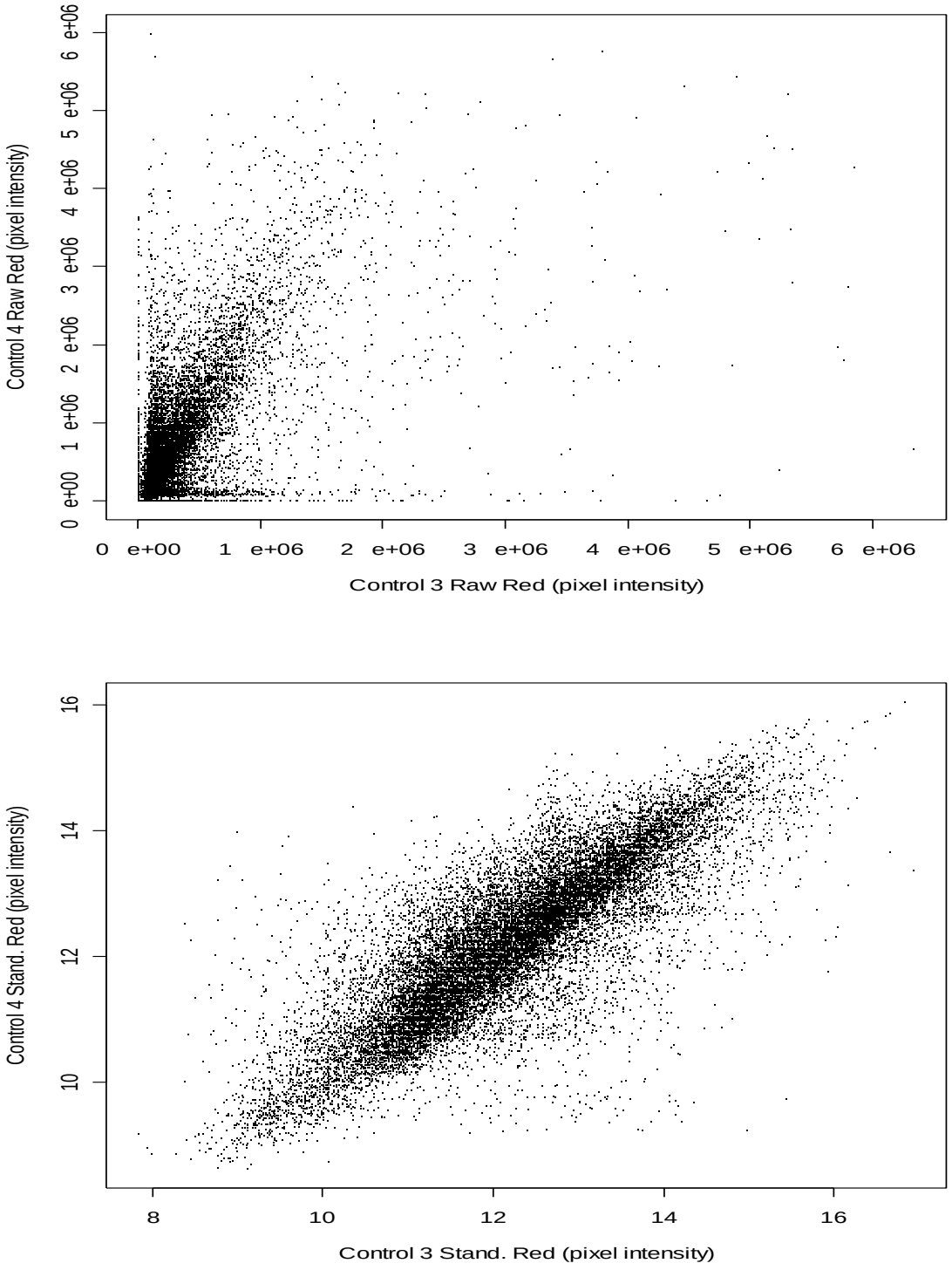

Figure 34. Scatterplot of Red Intensities for Dye Duplicates in Microarrays 7 and 8. For microarrays 7 and 8, both control samples were labeled with Cy5 (red) fluorescent dye. In this scatterplot, both pixel intensities were compared to ensure uniform and effective labeling. Pixel intensity equals the amount of signal from gene hybridization.

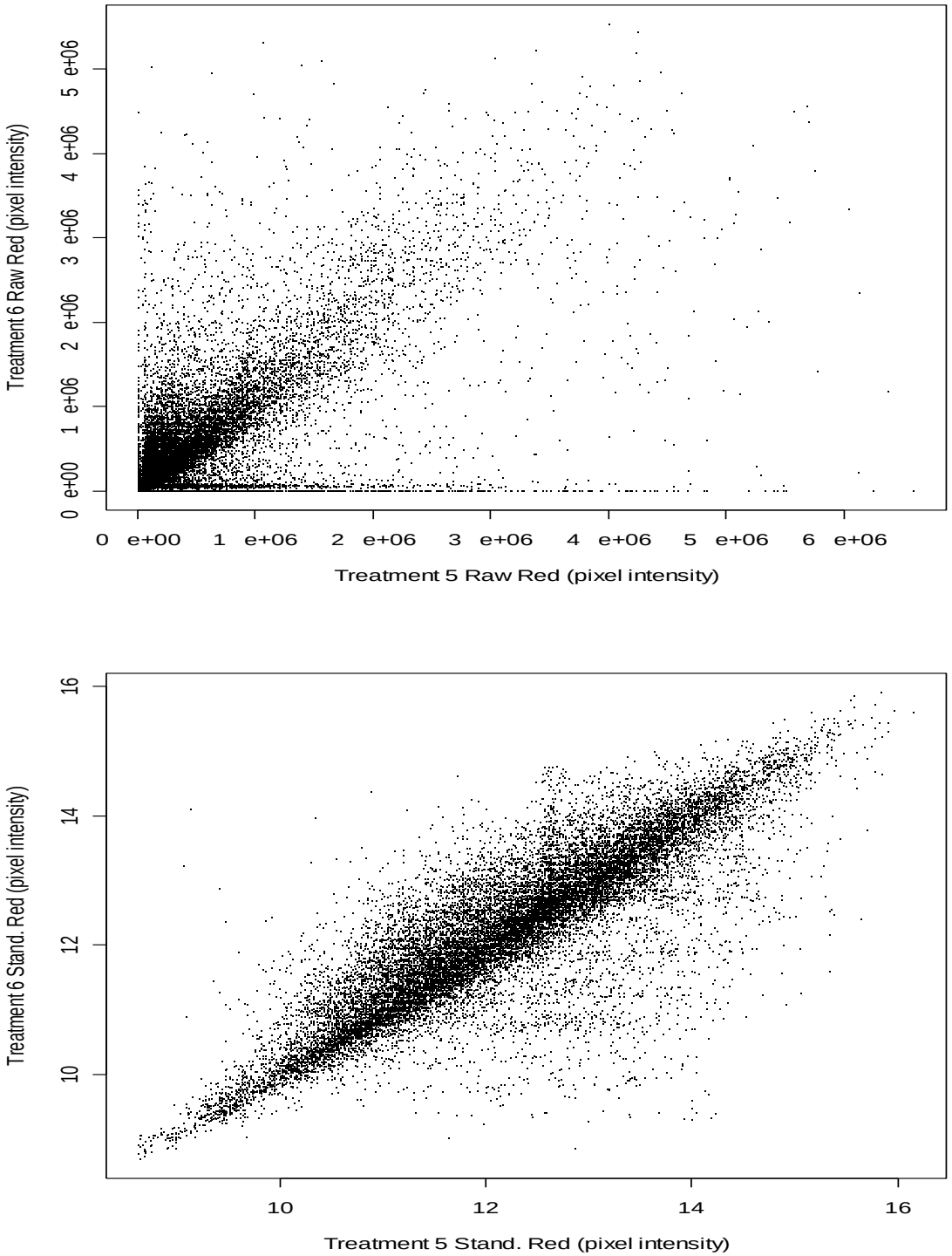

Figure 35. Scatterplot of Red Intensities for Dye Duplicates in Microarrays 9 and 10. For microarrays 9 and 10, both treatment samples were labeled with Cy5 (red) fluorescent dye. In this scatterplot, both pixel intensities were compared to ensure uniform and effective labeling. Pixel intensity equals the amount of signal from gene hybridization.

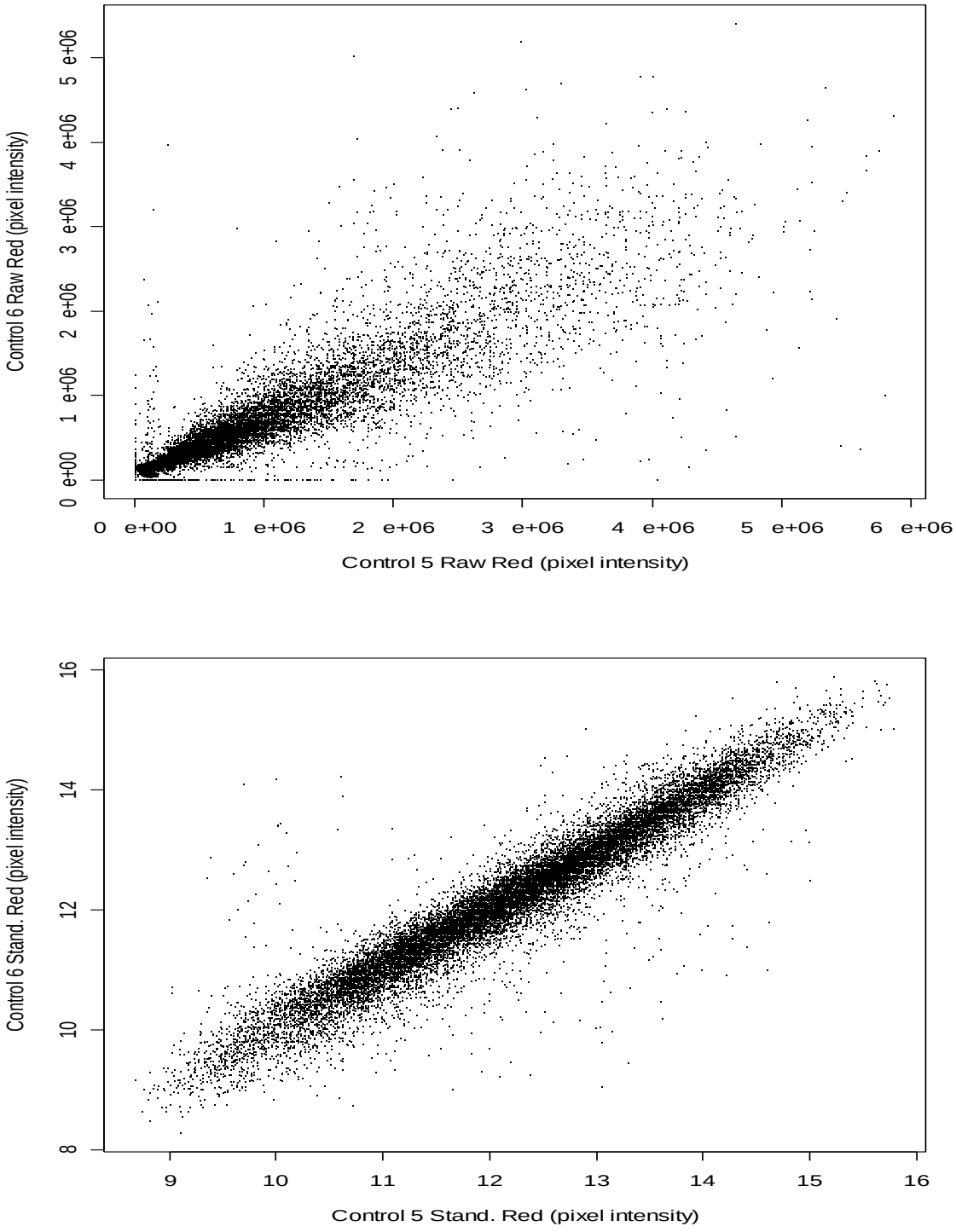

Figure 36. Scatterplot of Red Intensities for Dye Duplicates in Microarrays 11 and 12. For microarrays 11 and 12, both control samples were labeled with Cy5 (red) fluorescent dye. In this scatterplot, both pixel intensities were compared to ensure uniform and effective labeling. Pixel intensity equals the amount of signal from gene hybridization.

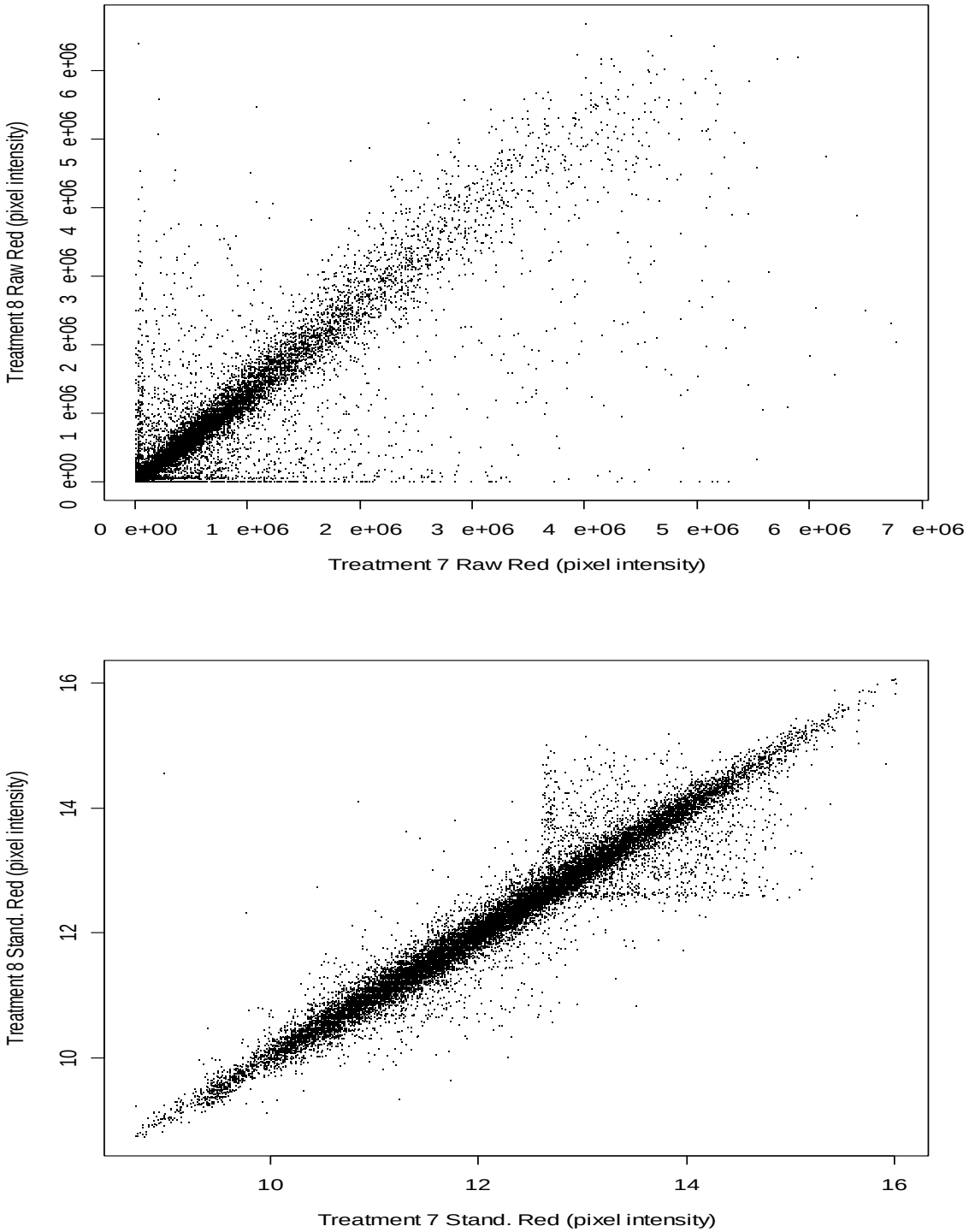

Figure 37. Scatterplot of Red Intensities for Dye Duplicates in Microarrays 13 and 14. For microarrays 13 and 14, both treatment samples were labeled with Cy5 (red) fluorescent dye. In this scatterplot, both pixel intensities were compared to ensure uniform and effective labeling. Pixel intensity equals the amount of signal from gene hybridization.

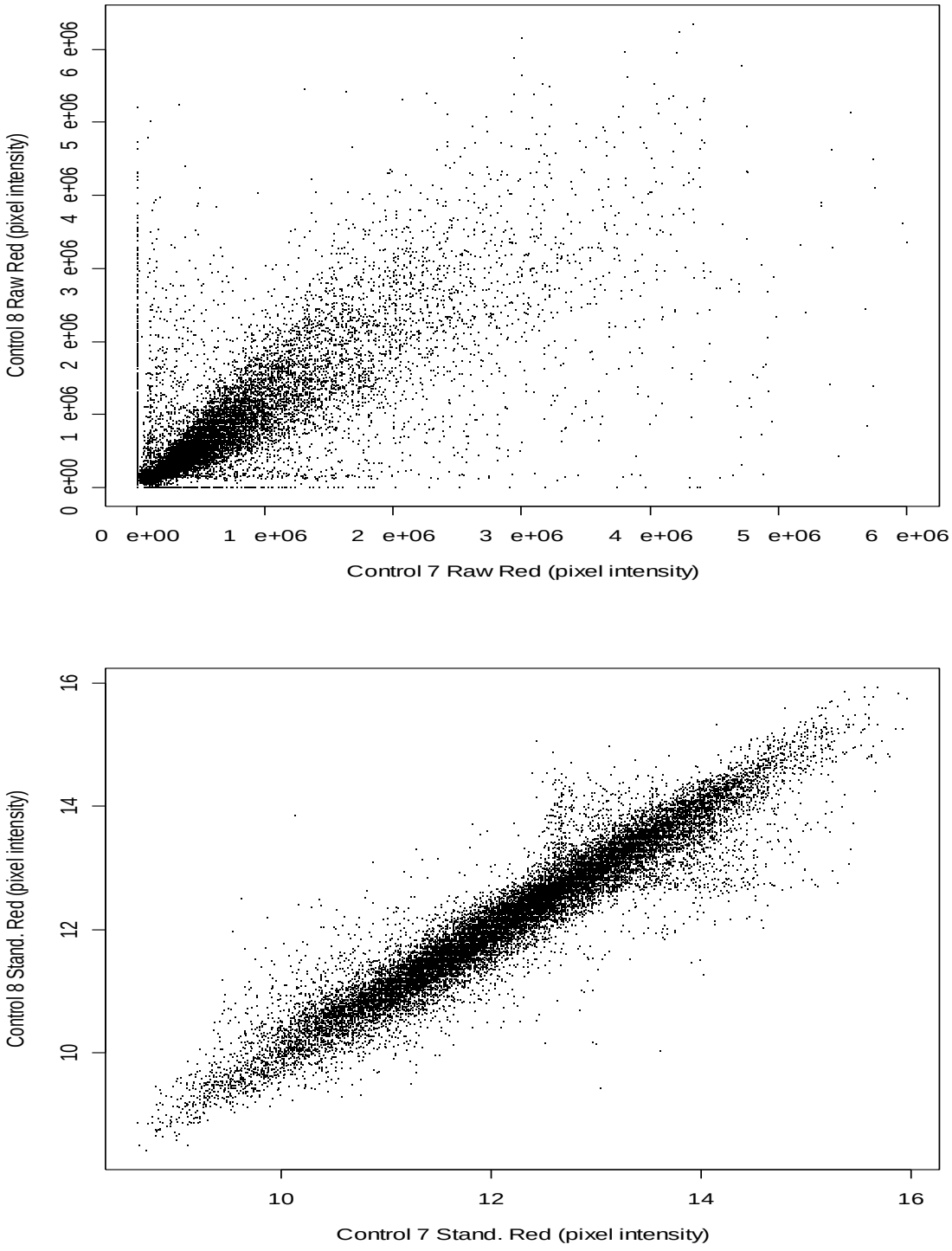

Figure 38. Scatterplot of Red Intensities for Dye Duplicates in Microarrays 15 and 16. For microarrays 15 and 16, both control samples were labeled with Cy5 (red) fluorescent dye. In this scatterplot, both pixel intensities were compared to ensure uniform and effective labeling. Pixel intensity equals the amount of signal from gene hybridization.

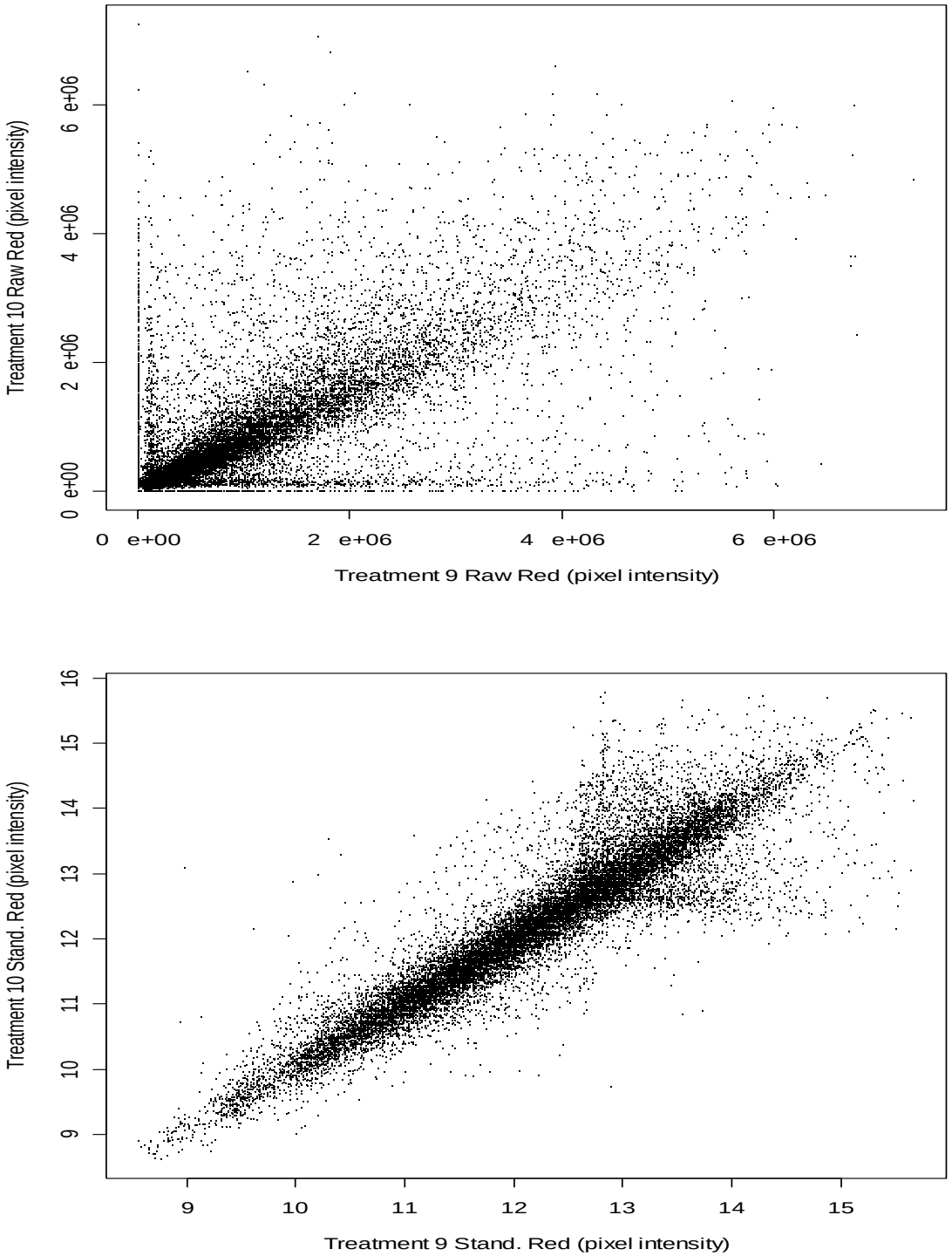

Figure 39. Scatterplot of Red Intensities for Dye Duplicates in Microarrays 17 and 18. For microarrays 17 and 18, both treatment samples were labeled with Cy5 (red) fluorescent dye. In this scatterplot, both pixel intensities were compared to ensure uniform and effective labeling. Pixel intensity equals the amount of signal from gene hybridization.

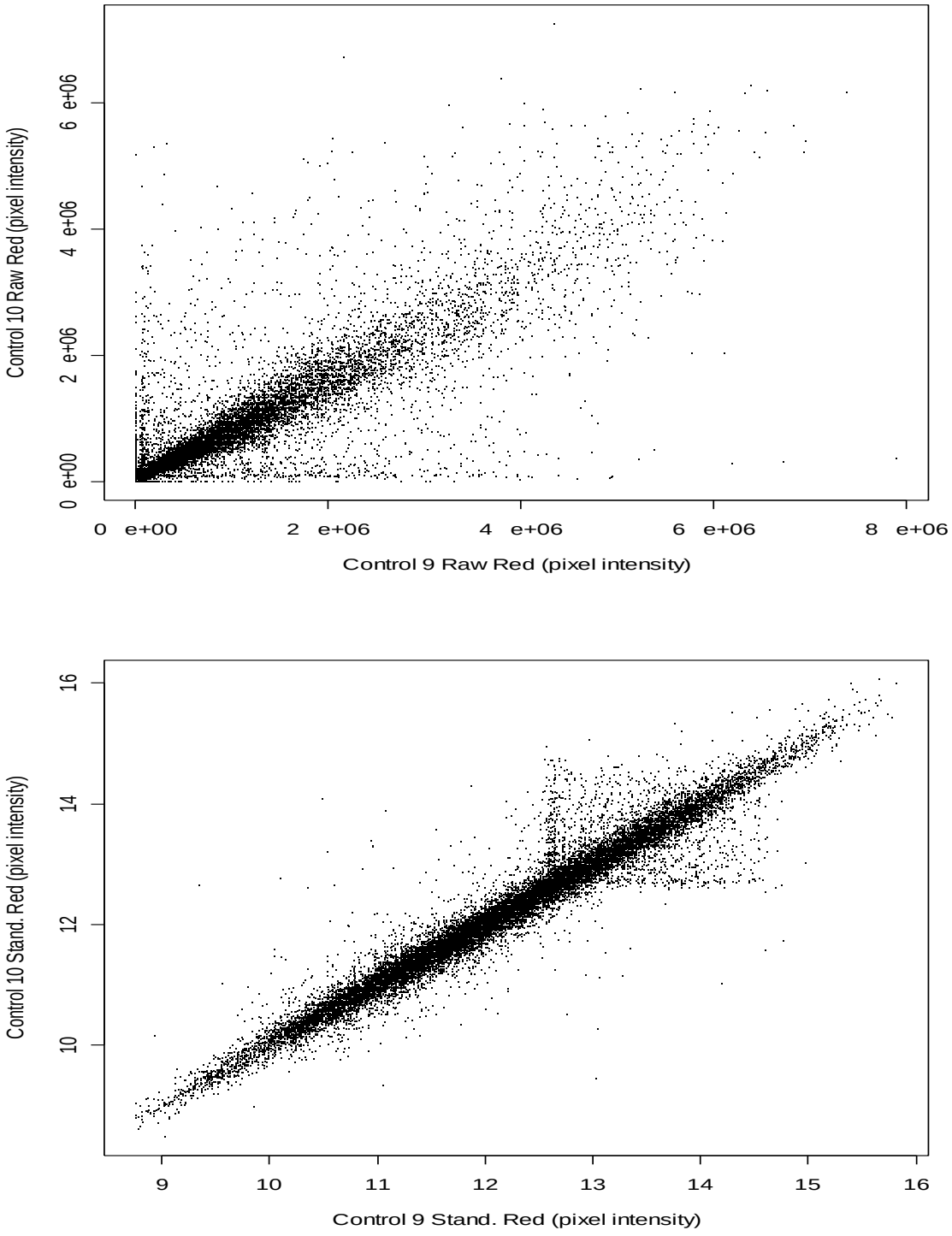

Figure 40. Scatterplot of Red Intensities for Dye Duplicates in Microarrays 19 and 20. For microarrays 19 and 20, both control samples were labeled with Cy5 (red) fluorescent dye. In this scatterplot, both pixel intensities were compared to ensure uniform and effective labeling. Pixel intensity equals the amount of signal from gene hybridization.

Significant Expression Differences and Similarities Resulting from UV Treatment

It was observed that each timepoint varied in the numbers of genes significantly affected by treatment. After 12 hr of UV exposure, the largest number of treatment-affected genes was evident. This was followed by 8 hr, 2 hr, 1 hr, and 4 hr (Table 1a). It was also observed that many mutually exclusive genes (only found at one timepoint) were present at each timepoint following the same pattern; after 12 hr the largest number of genes was evident followed by 8 hr, 2 hr, 1 hr, and 4 hr (Table 1b).

I identified overlap between the five significant gene lists. This resulted in 2-way, 3-way, 4-way and 5-way comparisons of the significant gene lists. The 2-way comparisons resulted in the largest total number of genes followed by 3-way, 4-way and 5-way comparisons (Table 2). Interestingly, when all five significant gene lists were compared to determine which, if any, genes were significantly differentially expressed across all timepoints, only 2 genes exhibited this phenomenon. They were AI674008 and BQ293537 (Table 2). No gene ontology information was available for gene BQ293537, however, three ontologies were associated with gene AI674008. They were as follows; intracellular protein transport, protein carrier activity, and membrane.

Table 1. Total Number of Significant Genes per Timepoint at 1% False Discovery Rate

Timepoint (hr of UV exposure)	Total Significant Genes
1	2314
2	4208
4	1783
8	4569
12	4735

(a) Total number of significant genes per timepoint whereby each list is not mutually exclusive.

Timepoint (hr of UV exposure)	Total Significant Genes
1	1009
2	1964
4	757
8	2128
12	2277

(b) Total number of significant genes found in specific timepoint and no other (mutually exclusive).

Table 2. Overlap in Significant Gene Lists (set intersections)

2-way	Total Same Genes	3-way	Total Same Genes	4-way	Total Same Genes	5-way	Total Same Genes
1,2	257	1,2,4	19	1,2,4,8	8	1,2,4,8,12	2
1,4	75	1,2,8	83	1,2,4,12	3		
1,8	283	1,2,12	61	1,2,8,12	24		
1,12	327	1,4,8	23	1,4,8,12	9		
2,4	209	1,4,12	29	2,4,8,12	22		
2,8	628	1,8,12	102				
2,12	622	2,4,8	66				
4, 8	213	2,4,12	56				
4, 12	223	2,8,12	184				
8, 12	25	4,8,12	69				

All possible comparisons between the five significant gene lists, where (1) corresponds to 1 hr of UV exposure, (2) corresponds to 2 hr of UV exposure, (4) corresponds to 4 hr of UV exposure, (8) corresponds to 8 hr of UV exposure, and (12) corresponds to 12 hr of UV exposure.

Expression Profiles of Treatment Affected Genes Across Time

The ORIOGEN program, based on order-restricted inference of intensity levels, was used to select genes for profiles according to their expression patterns over time. Treatment and control intensities were analyzed separately by ORIOGEN. Out of a total of approximately 19K genes, 348 genes were selected by the program in 1 of 8 control profiles, and 515 genes were selected in 1 of 8 treatment profiles. For the treatment profiles, the largest number of genes was selected in Profile 7 (Inverted Umbrella Profile, Upturn after 4 hr of UV) (Figure 47). This was followed by Profile 1, Increasing pattern of expression across time (Figure 41); Profile 2, Decreasing expression pattern over time (Figure 42); Profile 4 (Umbrella Profile, Downturn after 4 hr of UV) (Figure 44); Profile 5 (Umbrella Profile, Downturn after 8 hr of UV) (Figure 45); Profile 3 (Umbrella Profile, Downturn after 2 hr of UV) (Figure 43); Profile 8 (Inverted Umbrella Profile, Upturn after 8 hr of UV) (Figure 48); and Profile 6 (Inverted Umbrella Profile, Upturn after 2 hr of UV) (Figure 46). All genes selected are because of their fit to profiles, and we cannot determine if they are significantly different within a profile using ORIOGEN.

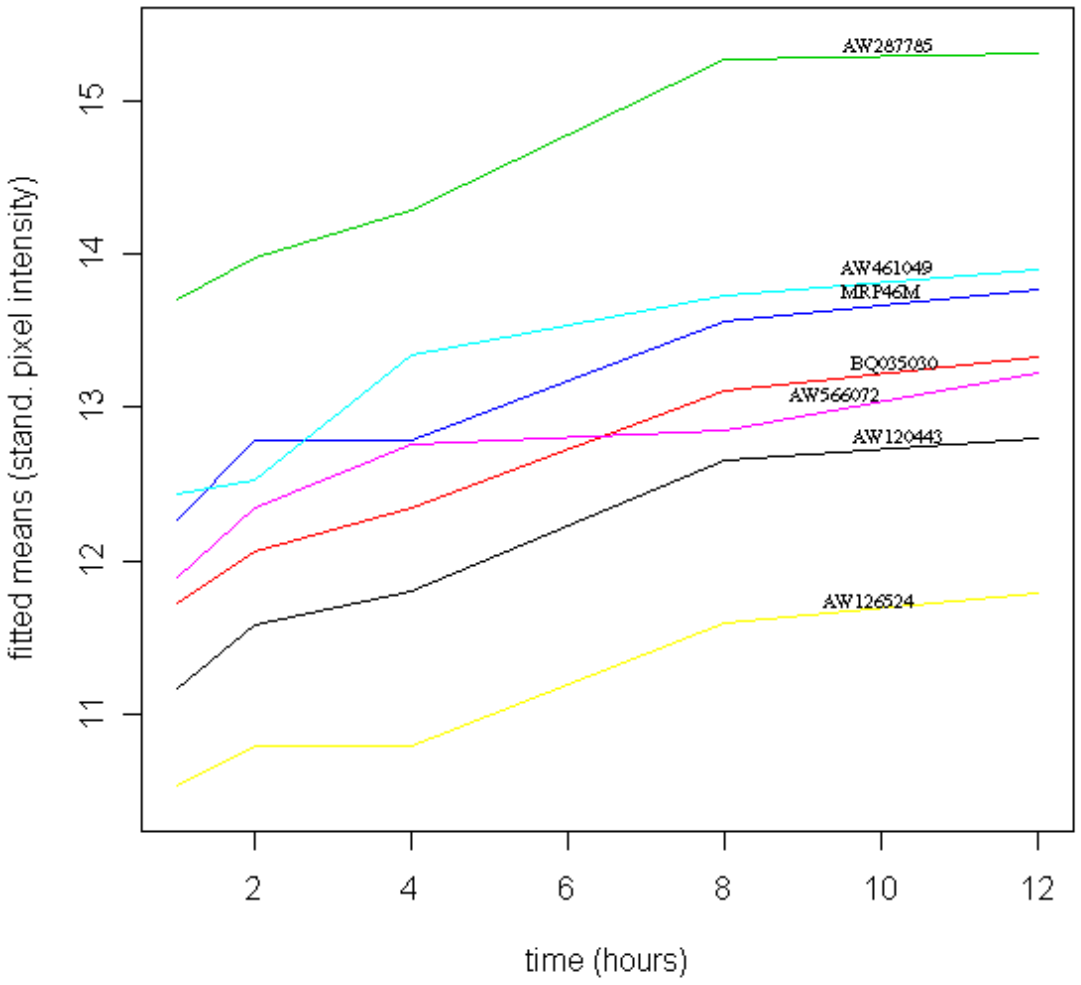

Figure 41. Treatment Profile 1 generated by ORIOGEN profile analysis software for genes from treatment samples that had a significant increase in mean expression over time. The seven genes with the largest fitted mean range (maximum-minimum) were selected for illustration. There were five timepoints utilized in this study corresponding to 1, 2, 4, 8, and 12 hr of UV exposure. Values 6 & 10 are for graphing purposes only.

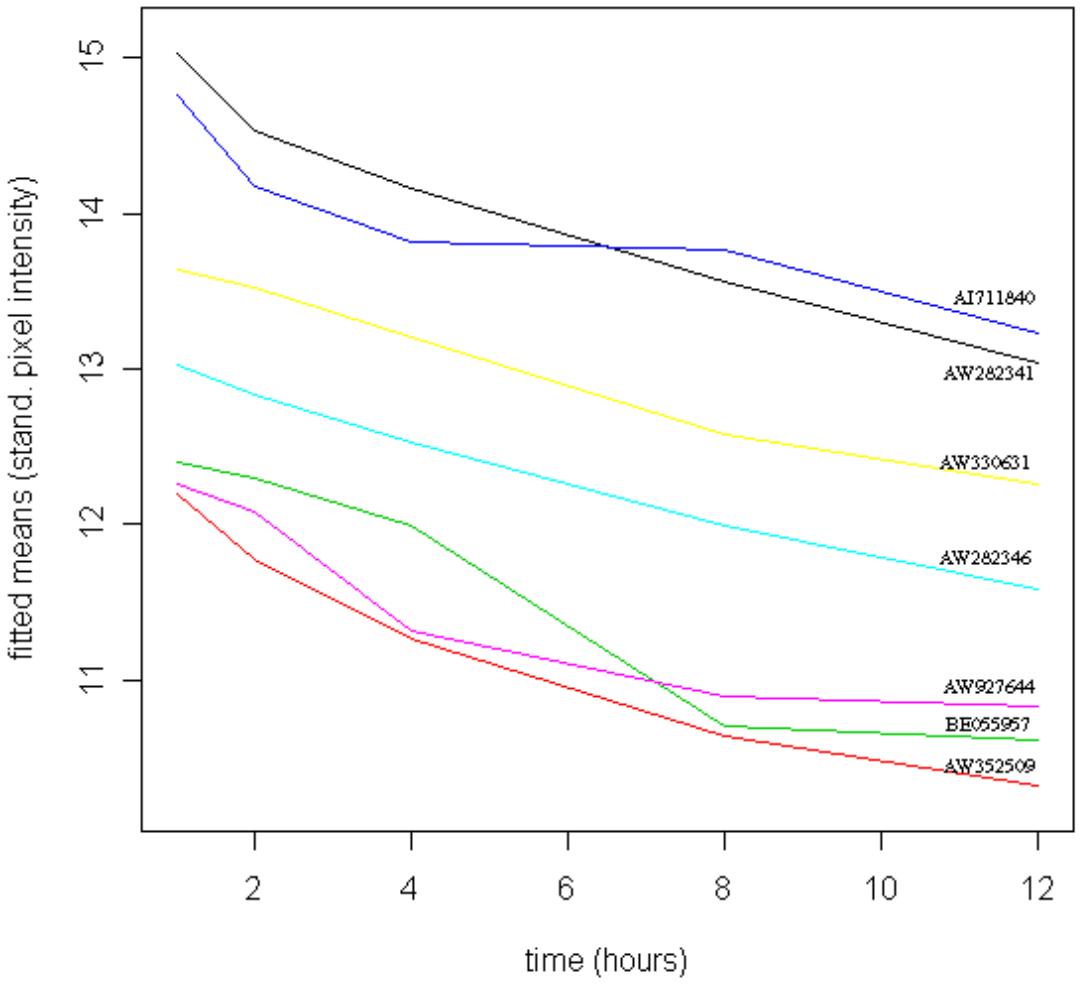

Figure 42. Treatment Profile 2 generated by ORIOGEN profile analysis software for genes from treatment samples that had a significant decrease in mean expression over time. The seven genes with the largest fitted mean range (maximum-minimum) were selected for illustration. There were five timepoints utilized in this study corresponding to 1, 2, 4, 8, and 12 hr of UV exposure. Values 6 & 10 are for graphing purposes only.

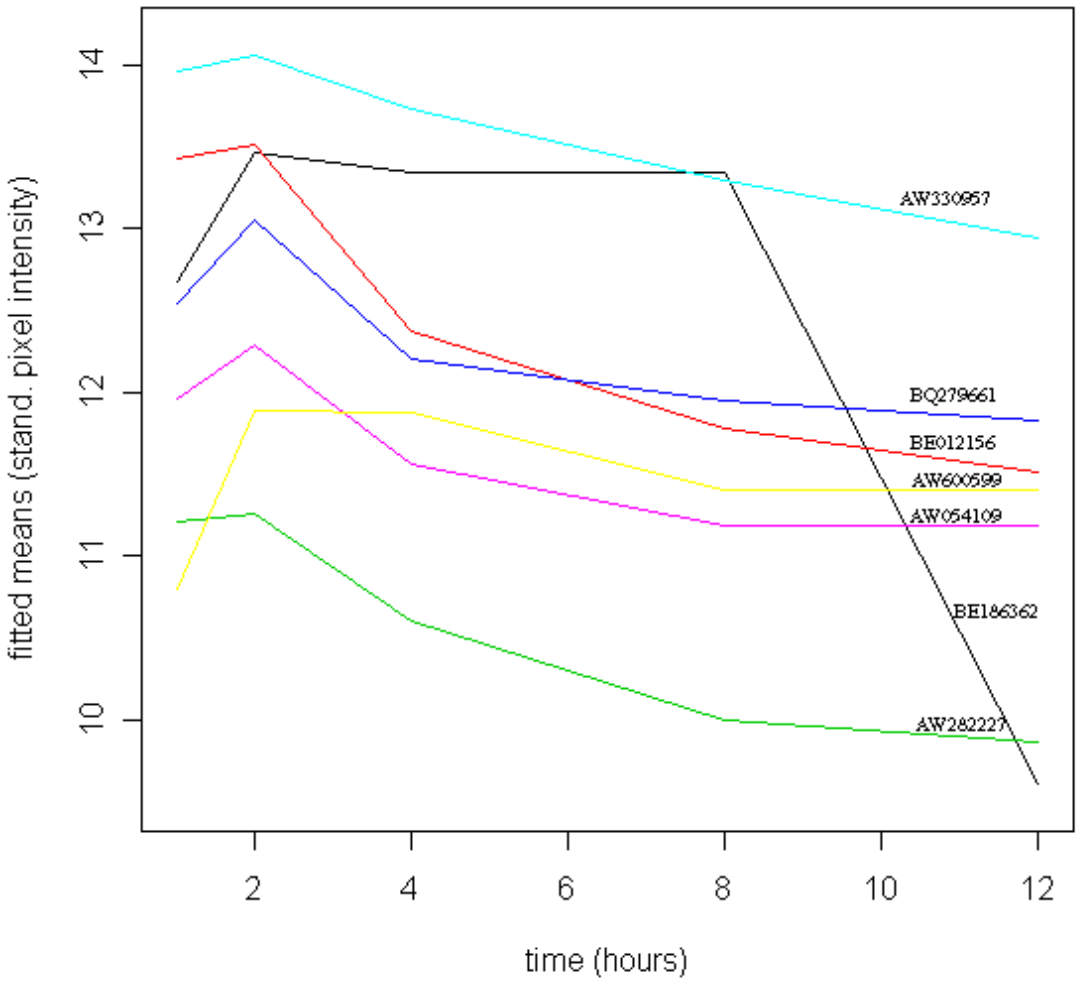

Figure 43. Treatment Profile 3 generated by ORIOGEN profile analysis software for genes from treatment samples that had a significant increase in mean expression with a downturn after 2 hr of UV exposure. The seven genes with the largest fitted mean range (maximum-minimum) were selected for illustration. There were five timepoints utilized in this study corresponding to 1, 2, 4, 8, and 12 hr of UV exposure. Values 6 & 10 are for graphing purposes only.

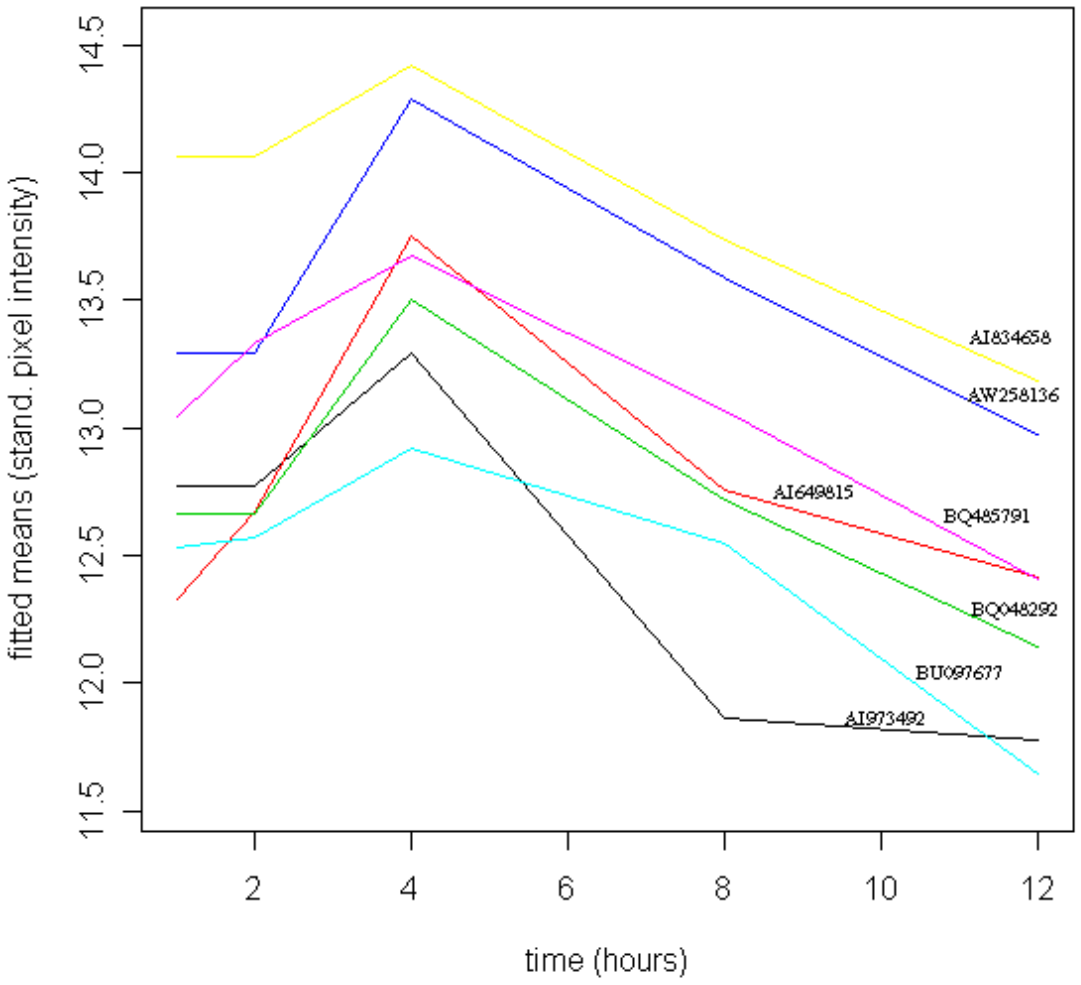

Figure 44. Treatment Profile 4 generated by ORIOGEN profile analysis software for genes from treatment samples that had a significant increase in mean expression with a downturn after 4 hr of UV exposure. The seven genes with the largest fitted mean range (maximum-minimum) were selected for illustration. There were five timepoints utilized in this study corresponding to 1, 2, 4, 8, and 12 hr of UV exposure. Values 6 & 10 are for graphing purposes only.

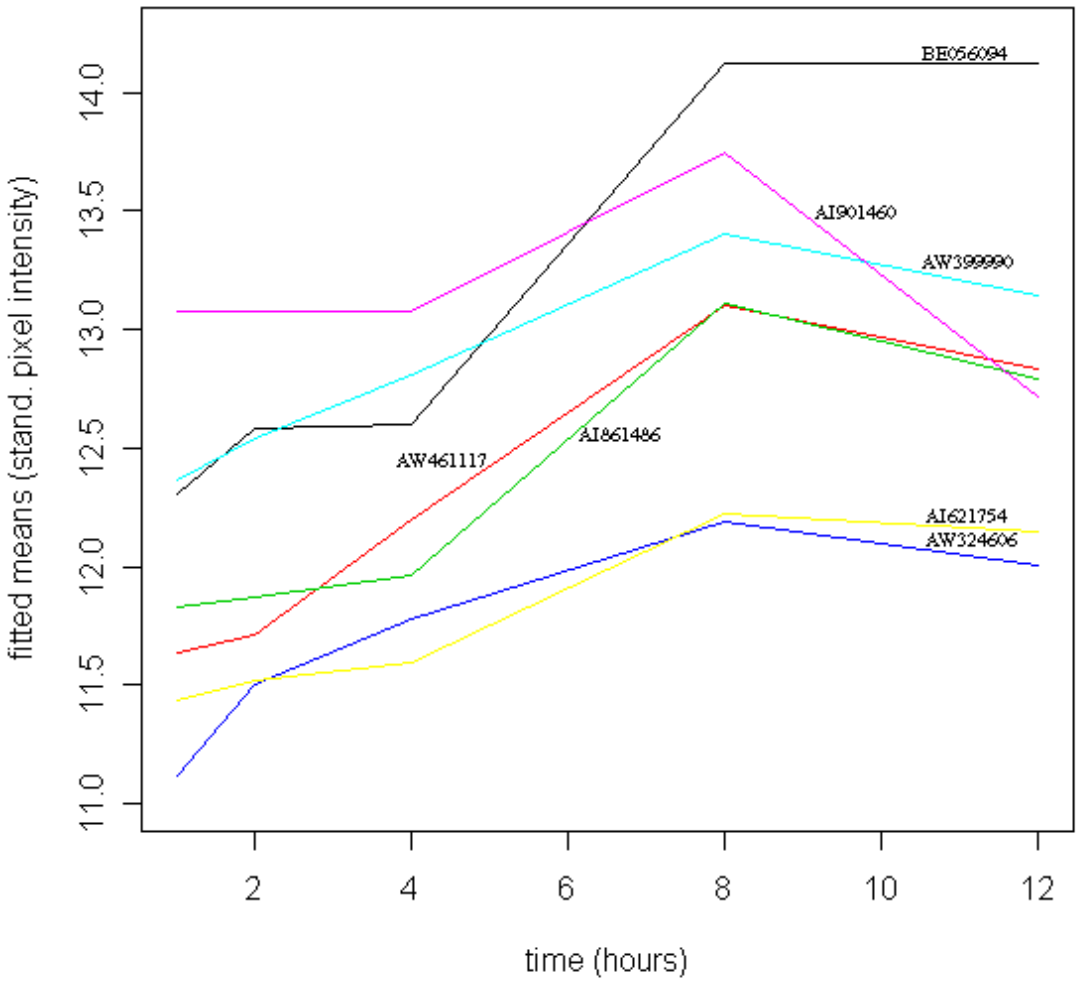

Figure 45. Treatment Profile 5 generated by ORIOGEN profile analysis software for genes from treatment samples that had a significant increase in mean expression with a downturn after 8 hr of UV exposure. The seven genes with the largest fitted mean range (maximum-minimum) were selected for illustration. There were five timepoints utilized in this study corresponding to 1, 2, 4, 8, and 12 hr of UV exposure. Values 6 & 10 are for graphing purposes only.

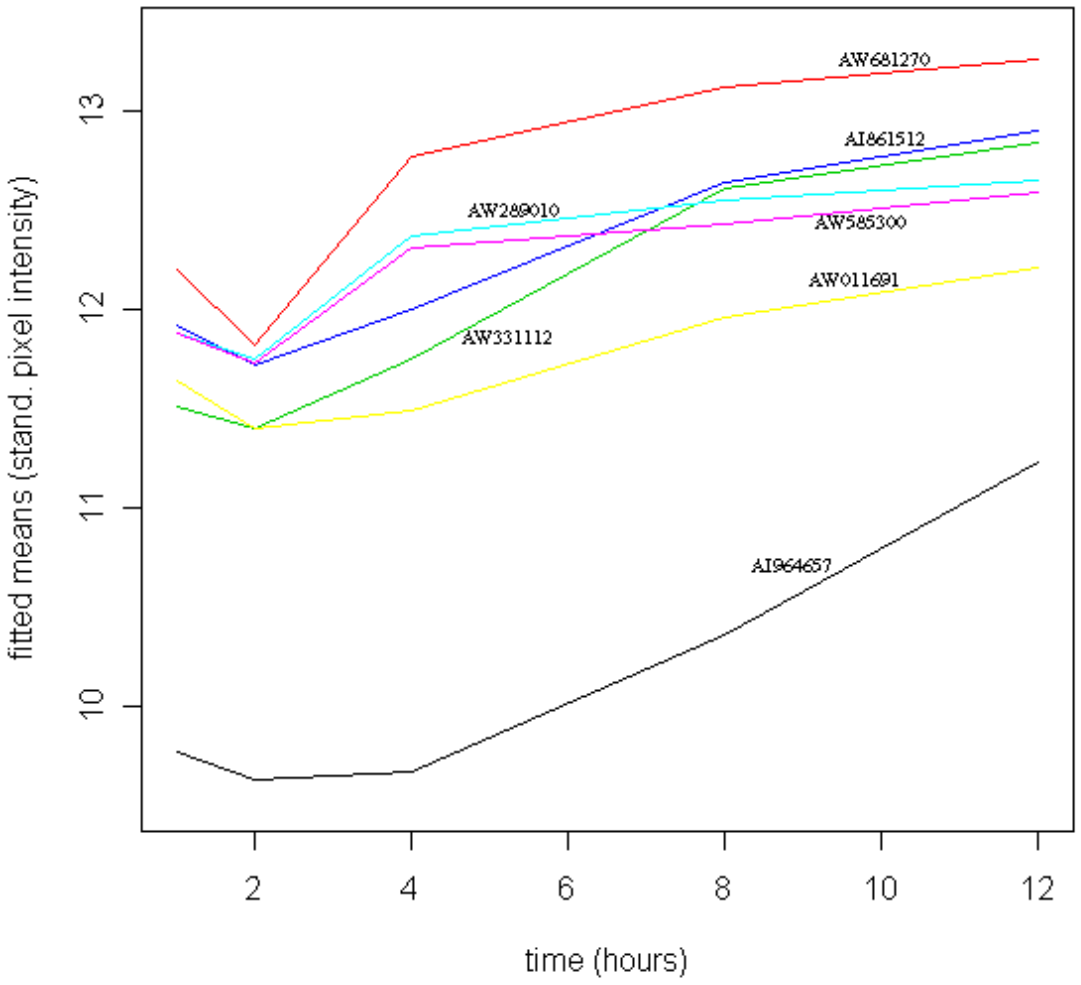

Figure 46. Treatment Profile 6 generated by ORIOGEN profile analysis software for genes from treatment samples that had a significant increase in mean expression with a upturn after 2 hr of UV exposure. The seven genes with the largest fitted mean range (maximum-minimum) were selected for illustration. There were five timepoints utilized in this study corresponding to 1, 2, 4, 8, and 12 hr of UV exposure. Values 6 & 10 are for graphing purposes only.

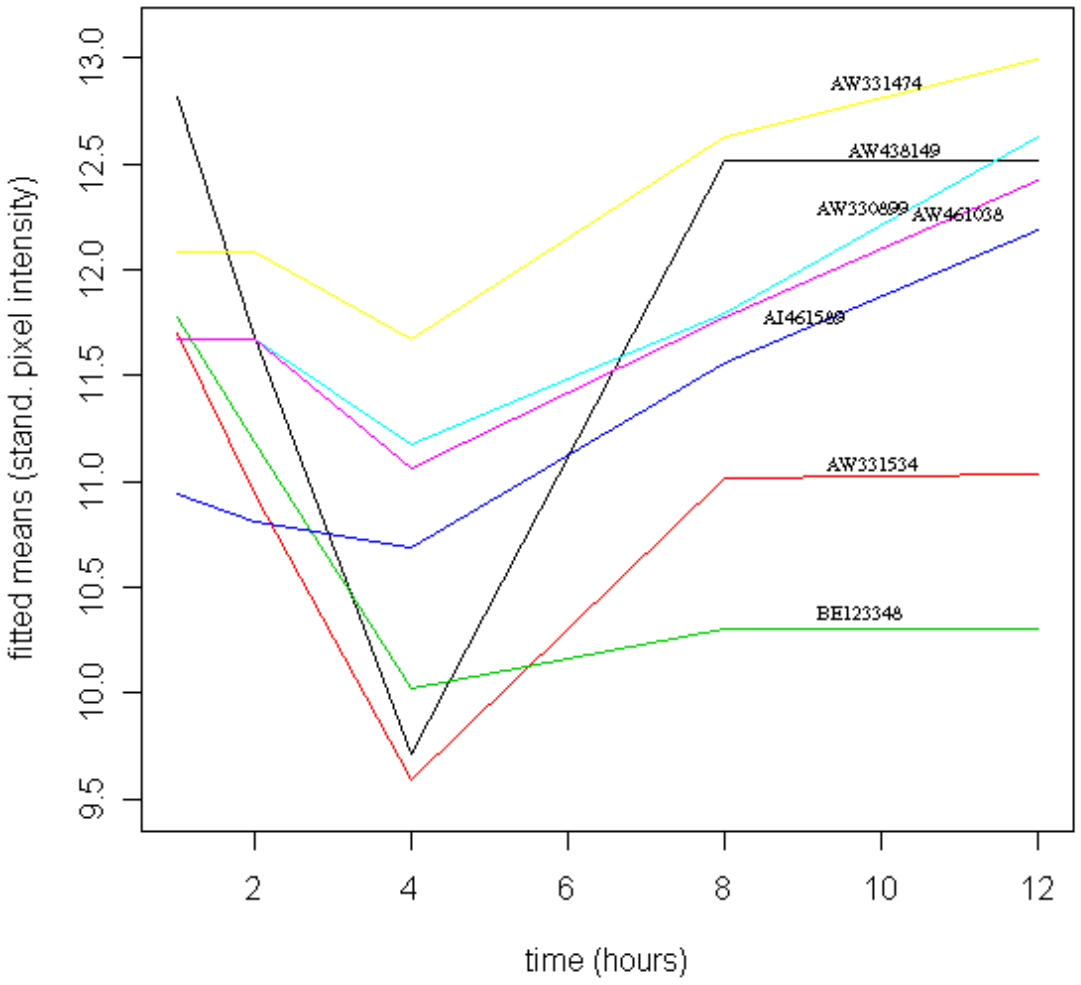

Figure 47. Treatment Profile 7 generated by ORIOGEN profile analysis software for genes from treatment samples that had a significant increase in mean expression with a upturn after 4 hr of UV exposure. The seven genes with the largest fitted mean range (maximum-minimum) were selected for illustration. There were five timepoints utilized in this study corresponding to 1, 2, 4, 8, and 12 hr of UV exposure. Values 6 & 10 are for graphing purposes only.

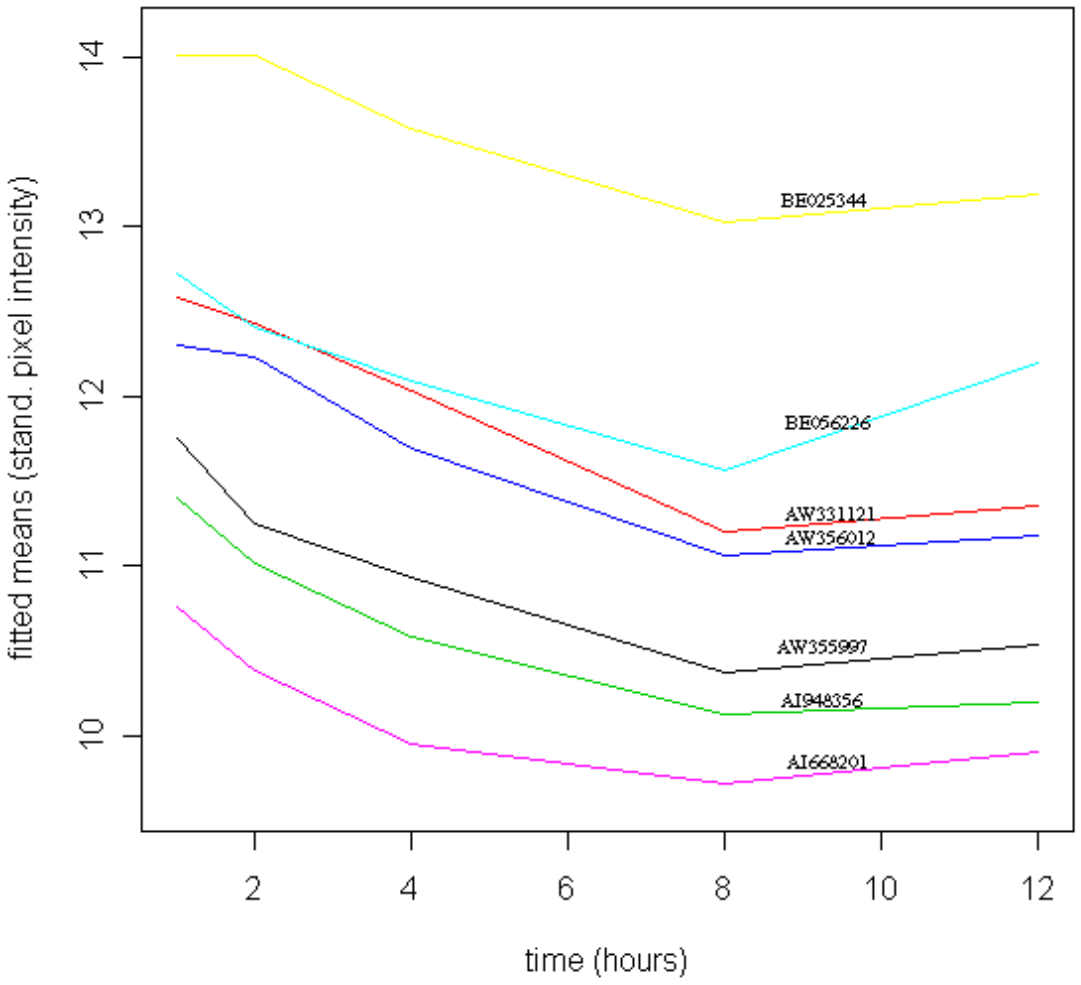

Figure 48. Treatment Profile 8 generated by ORIOGEN profile analysis software for genes from treatment samples that had a significant increase in mean expression with a upturn after 8 hr of UV exposure. The seven genes with the largest fitted mean range (maximum-minimum) were selected for illustration. There were five timepoints utilized in this study corresponding to 1, 2, 4, 8, and 12 hr of UV exposure. Values 6 & 10 are for graphing purposes only.

Gene Description Term Analysis

In addition to analyzing the gene ontology information for significant gene lists, I analyzed gene description terms. Of the 20232 genes on the microarray, 11779 had some type of text annotation available. ALACK, significance analysis of lexical bias in microarrays (Kim & Falkow, 2003), was utilized to analyze the gene description terms in order to identify significantly over-represented terms in each list. Gene description terms in each significant gene list, each mutually exclusive significant gene list, and each treatment profile list wee analyzed. As seen in Table 3, the significant gene lists yielded a variety of over-represented terms examples of which include "adenylate" P=.97 and "stearoyl" P=.98 (Table 3a). The mutually exclusive gene list for 1 hr UV yielded fewer significant terms, some of which were "ATP" with P=.95, "glutathione" with P=.94" and "vacuolar" with P=.99 (Table 3b). Significant gene list after 2 hr of UV yielded terms such as "acyl" with P=.95 and "cDNA" with P=.98 (Table 4a). The mutually exclusive gene list for 2 hr UV yielded significant terms such as "methyltransferase" with P=.987 and "xyloglucan" with P=.96 (Table 4b). Overall, terms identified as significantly over-represented were most numerous in inclusive significant gene lists and less numerous in the mutually exclusive significant gene lists (Tables 5-7). Treatment profile gene lists were also analyzed using ALACK to identify significantly over-represented gene description terms. Out of a total of eight treatment profile gene lists analyzed, one, Treatment profile 7, Inverted umbrella profile, upturn at 4 hours of UV (Table 8), was identified to have significantly over-represented gene description terms. In this list, the only term identified as significantly over-represented was "DNA" with P=.97. Protein

was identified as significantly over-represented but ALACK was unable to generate a p-value for it.

Table 3. Significantly Over-Represented Gene Description Terms in 1-hr Significant Gene List

Term	P value	% Occurrence
adenylate	0.97	
aminopeptidase*	0.91	33 (3/9)
ATPase	0.95	
beta	0.95	
carrier	0.93	
chain	0.92	
coat	1.00	
coatmer	0.97	
dependent	0.90	
glucosidase	0.94	
kaurene	1.00	
kinase	0.97	
membrane	0.90	
phosphoenolpyruvate*	0.94	22 (4/18)
precursor	0.91	
profilin	0.98	
proteasome	0.92	
recognition	0.92	
repair	0.90	
signal	0.98	
stearoyl	0.98	

(a) All significantly over-represented Homolog 1 Gene Descriptions from treatment-affected genes found in 1 hr significant gene list. Percent Occurrence equals (the total number of times the terms appears in the specific list divided by the total number of times the term occurs in the complete list) quantity multiplied by 100.
* Indicates significantly over-represented terms in significant gene lists that were randomly selected to determine percent occurrence.

Term	P value	% Occurrence
60S	0.99	
ATP	0.95	
glutathione*	0.94	9 (3/33)
H4	0.97	
helicase	0.96	
histone	0.90	
methyltransferase*	0.95	11 (3/27)
polyprotein	0.91	
putative	0.92	
receptor	0.91	
ribosomal	1.00	
threonine	0.94	
transferase	0.97	
UDP	0.94	
vacuolar	0.99	

(b) All significantly over-represented Homolog 1 Gene Descriptions from treatment-affected genes found in 1 hr significant gene list exclusively. Percent Occurrence equals (the total number of times the terms appears in the specific list divided by the total number of times the term occurs in the complete list) quantity multiplied by 100.
* Indicates significantly over-represented terms in significant gene lists that were randomly selected to determine percent occurrence.

Table 4. Significantly Over Represented Gene Description Terms in 2-hr Significant Gene List

Term	P value	% Occurrence
acyl	0.95	
apyrase	0.98	
ATPase	0.96	
carrier	0.95	
cDNA	0.98	
chain	0.92	
coat	0.94	
hydroxymethylglutaryl*	0.96	67 (4/6)
kaurene	0.97	
kinesin	0.98	
meth	0.90	
pectin	0.98	
pectinesterase*	0.98	80 (4/5)
pollen	0.96	
profilin	0.94	
proteasome	0.92	
RNase	0.91	
signal	0.96	
signalling	0.93	
stearoyl	0.95	
transposase	0.94	
vacuolar	0.92	

(a) All significantly over-represented Homolog 1 Gene Descriptions from treatment-affected genes found in 2 hr significant gene list. Percent Occurrence equals (the total number of times the terms appears in the specific list divided by the total number of times the term occurs in the complete list) quantity multiplied by 100.
* Indicates significantly over-represented terms in significant gene lists that were randomly selected to determine percent occurrence.

Term	P value	% Occurrence
40S	0.94	
ankyrin	0.90	
copper	0.97	
cysteinyl	0.99	
ferredoxin	0.93	
glucosidase*	0.96	20 (5/25)
malate	0.90	
methyltransferase	0.98	
myosin5	0.96	
phosphoglycerate*	0.91	18 (3/17)
plastid	0.92	
polyamine	1.00	
ribosomal	1.00	
synthetase	0.94	
transferase	0.94	
UDP	0.92	
vacuolar	0.91	
xyloglucan	0.96	

(b) All significantly over-represented Homolog 1 Gene Descriptions from treatment-affected genes found in 2 hr significant gene list only. Percent Occurrence equals (the total number of times the terms appears in the specific list divided by the total number of times the term occurs in the complete list) quantity multiplied by 100.
* Indicates significantly over-represented terms in significant gene lists that were randomly selected to determine percent occurrence.

Table 5. Significantly Over-Represented Gene Description Terms in 4-hr Significant Gene List

Term	P value	% Occurrence
ATP	0.95	
ATPase*	0.96	16 (10/63)
beta	0.90	
carrier	0.99	
chain	0.91	
coatmer	0.98	
dependent	0.90	
glucosidase	0.94	
inducible	0.90	
kinase	0.95	
membrane	0.96	
profiling	0.99	
pyrophosphate*	0.92	19 (5/26)
recognition	0.96	
signal	0.98	
stearoyl	0.99	

(a) All significantly over-represented Homolog 1 Gene Descriptions from treatment-affected genes found in 4 hr significant gene list. Percent Occurrence equals (the total number of times the terms appears in the specific list divided by the total number of times the term occurs in the complete list) quantity multiplied by 100.

* Indicates significantly over-represented terms in significant gene lists that were randomly selected to determine percent occurrence.

Term	P value	% Occurrence
60S	1.00	
H4	0.99	
histone*	0.98	5 (6/120)
precursor	0.90	
ribosomal*	1.00	5 (17/340)
UDP	0.97	

(b) All significantly over-represented Homolog 1 Gene Descriptions from treatment-affected genes found in 4 hr significant gene list only. Percent Occurrence equals (the total number of times the terms appears in the specific list divided by the total number of times the term occurs in the complete list) quantity multiplied by 100.
* Indicates significantly over-represented terms in significant gene lists that were randomly selected to determine percent occurrence.

Table 6. Significantly Over-Represented Gene Description Terms in 8-hr Significant Gene List

Term	P value	% Occurrence
acyl	0.94	
apyrase	0.97	
ATPase	0.96	
cDNA	0.98	
chain	0.92	
farnesyl	0.97	
hydroxymethylglutaryl	0.96	
kaurene	0.96	
kinesin	0.98	
methylesterase*	0.90	41 (7/17)
MSP1	0.94	
pectin	0.99	
pectinesterase*	1.00	100 (5/5)
pollen	0.97	
profilin	0.91	
proteasome	0.91	
repair	0.93	
RNase	0.90	
signal	0.98	
signalling	0.98	
stearoyl	0.98	
synthetase	0.92	
transposase	0.92	
ubiquinol	0.90	
vacuolar	0.91	

(b) All significantly over-represented Homolog 1 Gene Descriptions from treatment-affected genes found in 8 hr significant gene list. Percent Occurrence equals (the total number of times the terms appears in the specific list divided by the total number of times the term occurs in the complete list) quantity multiplied by 100.
* Indicates significantly over-represented terms in significant gene lists that were randomly selected to determine percent occurrence.

Term	P value	% Occurrence
40S	0.97	
copper	0.97	
cysteinyl	0.99	
cytoplasmic*	0.95	22 (4/18)
glucosidase	0.93	
histone	0.98	
methyltransferase	0.97	
myosin	0.93	
phosphoglycerate	0.96	
plastid	0.90	
polyamine	1.00	
ribosomal	1.00	
tubulin	0.92	
UDP	0.95	
xyloglucan*	0.99	36 (4/11)

(b) All significantly over-represented Homolog 1 Gene Descriptions from treatment-affected genes found in 8 hr significant gene list only. Percent Occurrence equals (the total number of times the terms appears in the specific list divided by the total number of times the term occurs in the complete list) quantity multiplied by 100.
* Indicates significantly over-represented terms in significant gene lists that were randomly selected to determine percent occurrence.

Table 7. Significantly Over-Represented Gene Description Terms in 12-hr Significant Gene List

Term	P value	% Occurrence
acyl	0.95	
apyrase	0.97	
ATPase	0.96	
carrier	0.97	
cDNA	0.97	
coat	0.93	
farnesyl	0.96	
hydroxymethylglutaryl	0.95	
kaurene	0.96	
kinesin	0.97	
mevalonate	0.94	
Nonclathrin	0.93	
pectin	0.99	
pectinesterase	0.99	
pollen	0.96	
profilin	0.90	
repair	0.92	
Ser	0.96	
signal	0.97	
signalling	0.97	
stearoyl*	0.98	67 (6/9)
transposase*	0.91	44 (7/16)

(a) All significantly over-represented Homolog 1 Gene Descriptions from treatment-affected genes found in 12 hr significant gene list. Percent Occurrence equals (the total number of times the terms appears in the specific list divided by the total number of times the term occurs in the complete list) quantity multiplied by 100.
* Indicates significantly over-represented terms in significant gene lists that were randomly selected to determine percent occurrence.

Term	P value	% Occurrence
copper	0.94	
cysteinyl*	0.99	60 (3/5)
cytoplasmic	0.93	
glucosidase	0.90	
histone	0.98	
methyltransferase	0.95	
myosin	0.90	
phosphoglycerate	0.99	
polyamine	1.00	
ribosomal*	1.00	16 (59/369)
UDP	0.93	
vacuolar	0.91	
xyloglucan	0.98	

(b) All significantly over-represented Homolog 1 Gene Descriptions from treatment-affected genes found in 12 hr significant gene list only. Percent Occurrence equals (the total number of times the terms appears in the specific list divided by the total number of times the term occurs in the complete list) quantity multiplied by 100.
* Indicates significantly over-represented terms in significant gene lists that were randomly selected to determine percent occurrence.

Table 8. Significantly Over-Represented Gene Description Terms in Profile 7

Term	P value	% Occurrence
DNA*	0.97	2 (3/150)

All significantly over-represented Homolog 1 Gene Descriptions from treatment-affected genes found in Profile 7 (Inverted Umbrella Profile, Upturn at 3). Percent Occurrence equals (the total number of times the terms appears in the specific list divided by the total number of times the term occurs in the complete list) quantity multiplied by 100.
* Indicates significantly over-represented terms in significant gene lists that were randomly selected to determine percent occurrence.

Gene Ontology (GOID) Analysis

Structured gene description terms are available for Maize in the TIGR maize gene indices database (Ashburner & Lewis, 2002). All available gene ontology identification numbers (GOID's) were extracted for the genes on the microarray totaling approximately 33% of the total genes. GOID's were merged with each significant gene list and analyzed using chi-square frequency analysis (Beissbarth & Speed, 2004). Two lists were found to have over/under-represented GOID's; treatment affected genes after 1 hr of UV exposure (Table 9) and treatment affected genes after 4 hr (Table 10) of UV exposure. There were 53 over/under-represented GOID's identified in the 1 hr significant gene list with 77% being unique in comparison to the over/under-represented GOID's present in the 4 hr significant gene list. Examples of unique over/under-represented gene ontologies present in the 1 hr significant gene list included genes involved with "oxygen and reactive oxygen species metabolism, the chemical reactions involving dioxygen, or any of the reactive oxygen species, e.g. superoxide anions, hydrogen peroxide, and hydroxyl radicals", "cell communication, any process that mediates interactions between a cell and its surroundings, encompasses interactions such as signaling or attachment between one cell and another cell, between a cell and an extracellular matrix, or between a cell and any other aspect of its environment." The remaining 23% of over/under-represented GOID's present in the 1 hr significant gene list were shared with the 4 hr over/under-represented GOID list. Examples of these included, "processes interacting selectively with tRNA," "cytoplasmic mRNA processing," "protein biosynthesis," and "protein-nucleus export." There were 40 over/under-represented GOID's present in the 4 hr significant gene list with 58% being unique when

compared to the 1 hr significant gene list. Examples of over/under-represented gene ontologies unique to the 4 hr significant gene list are "mRNA catabolism, the nonsense mediated mRNA decay pathway that degrades mRNAs transcribed from genes in which an amino-acid codon has changed to a nonsense codon; this prevents the translation of such mRNAs into truncated, and potentially harmful, proteins," "NAPDH regeneration", "smoothened signaling pathway, the series of molecular signals generated as a consequence of activation of the transmembrane protein", "mRNA metabolism, the chemical reactions involving mRNA, messenger RNA, which is responsible for carrying the coded genetic 'message', transcribed from DNA, to sites of protein assembly at the ribosomes."

Each of eight treatment profiles were also analyzed to determine if over/under-represented GOID's existed. Four of eight profiles were found to have significantly over/under-represented GOID's: Treatment profile 2, Decreasing, (Table 11); Treatment profile 4, Umbrella profile, downturn at 4 hr of UV (Table 12); Treatment profile 5, Umbrella profile, downturn at 8 hr of UV (Table 13); and Treatment profile 7, Inverted umbrella profile, upturn at 4 hr of UV (Table 14). In Treatment profile 2, examples of over/under-represented GOID's include several with "heat shock protein activities, responses to heat, etc., processes involving specific groups of proteins that are synthesized in cells after being exposed to temperatures that are higher than normal or high light intensity," "mitochondrial electron transport, the transfer of electrons from ubiquinol to Cytochrome C that occurs during oxidative phosphorylation, mediated by the multisubunit enzyme known as complex III, and "the positive regulation of transcription." Examples of over/under-represented ontologies present in Treatment

profile 4 include, “cathrin adaptor complexes, and coats for vesicles,” “induction of cell apoptosis,” and “lipid transport.” Examples of over/under-represented GOID’s in Treatment profile 5 include, “nucleic acid binding, interacting selectively with any nucleic acid,” and “biosynthesis, the energy-requiring part of metabolism in which simpler substances are transformed into more complex ones, as in growth and other biosynthetic processes.” Examples of over/under-represented GOID’s in Treatment profile 7 include, “protein degradation tagging activity, covalent addition of polyubiquitin to another protein, targeting the tagged protein for destruction,” “carrier activity, catalysis of the transfer of a specific substance or related group of substances from one side of the membrane to the other,” and “mitochondrial electron transport, the transfer of electrons from NADH to ubiquinone that occurs during oxidative phosphorylation, mediated by the multisubunit enzyme known as complex I.”

Table 9. Significantly Over/Under Represented GOID's in 1 hr Significant Gene List for a 1% False Discovery Rate

GOID	Text
GO:0000049	tRNA binding,molecular_function Interacting selectively with transfer RNA.
GO:0000059	protein-nucleus import, docking, biological process
GO:0000280	nuclear division biological_process The partitioning of the nucleus and its genetic information.
GO:0000290	deadenylylation-dependent decapping biological_process Cleavage of the 5'-cap of an mRNA triggered by shortening of the poly(A) tail to below a minimum functional length.
GO:0000311	plastid large ribosomal subunit cellular_component The larger of the two subunits of a plastid ribosome. Two sites on the ribosomal large subunit are involved in translation: the aminoacyl site (A site) and peptidyl site (P site).
GO:0000932	cytoplasmic mRNA processing body cellular_component A focus in the cytoplasm where mRNAs may become inactivated by decapping or some other mechanism. mRNA processing and binding proteins are localized to these foci.
GO:0003735	structural constituent of ribosome molecular_function The action of a molecule that contributes to the structural integrity of the ribosome.
GO:0003862	3-isopropylmalate dehydrogenase activity molecular_function Catalysis of the reaction: 3-carboxy-2-hydroxy-4-methylpentanoate + NAD+ = 3-carboxy-4-methyl-2-oxopentanoate + NADH + H+.
GO:0003917	DNA topoisomerase type I activity molecular_function Catalysis of a DNA topological transformation by transiently cleaving one DNA strand at a time to allow passage of another strand; changes the linking number by +1 per catalytic cycle.
GO:0003929	RAN small monomeric GTPase activity molecular_function OBSOLETE. Catalysis of the reaction: GTP + H2O = GDP + phosphate.

GO:0003975	UDP-N-acetylglucosamine-dolichyl-phosphate N-acetylglucosaminephosphotransferase activity molecular_function Catalysis of the reaction: UDP-N-acetyl-D-glucosamine + dolichyl phosphate = UMP + N-acetyl-D-glucosaminyl-diphosphodolichol.
GO:0004396	hexokinase activity molecular_function Catalysis of the reaction: ATP + D-hexose = ADP + D-hexose 6-phosphate.
GO:0004514	nicotinate-nucleotide diphosphorylase (carboxylating) activity molecular_function Catalysis of the reaction: nicotinate D-ribonucleotide + diphosphate + CO2 = pyridine-2,3-dicarboxylate + 5-phospho-alpha-D-ribose 1-diphosphate.
GO:0004516	nicotinate phosphoribosyltransferase activity molecular_function Catalysis of the reaction: nicotinate D-ribonucleotide + diphosphate = nicotinate + 5-phospho-alpha-D-ribose 1-diphosphate.
GO:0004604	phosphoadenylyl-sulfate reductase (thioredoxin) activity molecular_function Catalysis of the reaction: adenosine 3',5'-bisphosphate + SO3(2-) (sulfite) + oxidized thioredoxin = 3'-phosphoadenosine 5'-phosphosulfate (PAPS) + reduced thioredoxin.
GO:0004630	phospholipase D activity molecular_function Catalysis of the reaction: a phosphatidylcholine + H2O = choline + a phosphatidate.
GO:0004637	phosphoribosylamine-glycine ligase activity molecular_function Catalysis of the reaction: ATP + 5-phospho-D-ribosylamine + glycine = ADP + phosphate + N1-(5-phospho-D-ribosyl)glycinamide.
GO:0004681	casein kinase I activity molecular_function
GO:0004815	aspartate-tRNA ligase activity molecular_function Catalysis of the reaction: ATP + L-aspartate + tRNA(Asp) = AMP + diphosphate + L-aspartyl-tRNA(Asp).
GO:0004824	lysine-tRNA ligase activity molecular_function Catalysis of the reaction: ATP + L-lysine + tRNA(Lys) =

	AMP + diphosphate + L-lysyl-tRNA(Lys).
GO:0005310	dicarboxylic acid transporter activity molecular_function Enables the directed movement of dicarboxylic acids into, out of, within or between cells. A dicarboxylic acid is an organic acid with two COOH groups.
GO:0005478	intracellular transporter activity molecular_function Enables the directed movement of substances within a cell.
GO:0005578	extracellular matrix (sensu Metazoa) cellular_component A layer consisting mainly of proteins (especially collagen) and glycosaminoglycans (mostly as proteoglycans) that forms a sheet underlying cells such as endothelial and epithelial cells. The proteins are secreted by cells in the vicinity. As in, but not restricted to, the multicellular animals (Metazoa, ncbi_taxonomy_id:33208).
GO:0005643	nuclear pore cellular_component Any of the numerous similar discrete openings in the nuclear envelope of a eukaryotic cell, where the inner and outer nuclear membranes are joined.
GO:0005794	Golgi apparatus cellular_component A compound membranous cytoplasmic organelle of eukaryotic cells, consisting of flattened, ribosome-free vesicles arranged in a more or less regular stack. The Golgi apparatus differs from the endoplasmic reticulum in often having slightly thicker membranes, appearing in sections as a characteristic shallow semicircle so that the convex side (cis or entry face) abuts the endoplasmic reticulum, secretory vesicles emerging from the concave side (trans or exit face). In vertebrate cells there is usually one such organelle, while in invertebrates and plants, where they are known usually as dictyosomes, there may be several scattered in the cytoplasm. The Golgi apparatus processes proteins produced on the ribosomes of the rough endoplasmic reticulum; such processing includes modification of the core

	oligosaccharides of glycoproteins, and the sorting and packaging of proteins for transport to a variety of cellular locations. Three different regions of the Golgi are now recognized both in terms of structure and function: cis, in the vicinity of the cis face, trans, in the vicinity of the trans face, and medial, lying between the cis and trans regions.
GO:0005843	cytosolic small ribosomal subunit (sensu Eukaryota) cellular_component The small subunit of a eukaryotic cytosolic ribosome; has a sedimentation coefficient of 40S. As in, but not restricted to, the eukaryotes (Eukaryota, ncbi_taxonomy_id:2759)
GO:0006323	DNA packaging biological_process Any process by which DNA and associated proteins are formed into a compact, orderly structure.
GO:0006412	protein biosynthesis biological_process The formation from simpler components of a protein, rather than of proteins in general.
GO:0006422	aspartyl-tRNA aminoacylation biological_process The process of coupling aspartate to aspartyl-tRNA, catalyzed by aspartyl-tRNA synthetase. In tRNA aminoacylation, the amino acid is first activated by linkage to AMP and then transferred to either the 2'- or the 3'-hydroxyl group of the 3'-adenosine residue of the tRNA.
GO:0006430	lysyl-tRNA aminoacylation biological_process The process of coupling lysine to lysyl-tRNA, catalyzed by lysyl-tRNA synthetase. In tRNA aminoacylation, the amino acid is first activated by linkage to AMP and then transferred to either the 2'- or the 3'-hydroxyl group of the 3'-adenosine residue of the tRNA.
GO:0006595	polyamine metabolism biological_process The chemical reactions involving polyamines, any organic compound containing two or more amino groups.
GO:0006611	protein-nucleus export biological_process The directed movement of a protein from the nucleus into the cytoplasm.

GO:0006800	oxygen and reactive oxygen species metabolism biological_process The chemical reactions involving dioxygen (O2), or any of the reactive oxygen species, e.g. superoxide anions (O2-), hydrogen peroxide (H2O2), and hydroxyl radicals (-OH).
GO:0006835	dicarboxylic acid transport biological_process The directed movement of dicarboxylic acids into, out of, within or between cells.
GO:0006840	mitochondrial alpha-ketoglutarate/malate transport biological_process OBSOLETE. The directed movement of alpha-ketoglutarate and malate into, out of or within a mitochondrion.
GO:0006880	intracellular sequestering of iron ion biological_process The process of binding or confining iron ions in an intracellular area such that they are separated from other components of a biological system.
GO:0006903	vesicle targeting biological_process Targeting of a vesicle to a specific destination membrane.
GO:0007052	mitotic spindle organization and biogenesis biological_process The formation and maintenance of the microtubule spindle during a mitotic cell cycle.
GO:0007154	cell communication biological_process Any process that mediates interactions between a cell and its surroundings. Encompasses interactions such as signaling or attachment between one cell and another cell, between a cell and an extracellular matrix, or between a cell and any other aspect of its environment.
GO:0007166	cell surface receptor linked signal transduction biological_process Any series of molecular signals initiated by the binding of an extracellular ligand to a receptor on the surface of the target cell.
GO:0007252	I-kappaB phosphorylation biological_process The process of introducing a phosphoric group into an I-kappaB protein.
GO:0008054	cyclin catabolism biological_process The

	breakdown into simpler components of cyclins, proteins whose levels in a cell varies markedly during the cell cycle, and which play key roles in regulating cell cycle phase transitions.
GO:0008379	thioredoxin peroxidase activity molecular_function Catalysis of the reaction: reduced thioredoxin + H2O2 = oxidized thioredoxin + H2O.
GO:0008430	selenium binding molecular_function Interacting selectively with selenium (Se).
GO:0008447	L-ascorbate oxidase activity molecular_function Catalysis of the reaction: 2 L-ascorbate + O2 = 2 dehydroascorbate + 2 H2O.
GO:0008462	endopeptidase Clp activity molecular_function Catalysis of the hydrolysis of proteins to small peptides in the presence of ATP and magnesium. Alpha-casein is the usual test substrate. In the absence of ATP, only oligopeptides shorter than five residues are cleaved, for example, succinyl-Leu-Tyr-NHMec which is cleaved at the Tyr-NHMec bond, and Leu-Tyr-Leu-Tyr-Trp which is cleaved at the second Leu-Typ bond (cleavage of the Tyr-Leu and Tyr-Trp bonds also occurs).
GO:0008785	alkyl hydroperoxide reductase activity molecular_function Catalysis of the reaction: octane hydroperoxide + NADH + H+ = H2O + NAD+ + 1-octanol.
GO:0009113	purine base biosynthesis biological_process The formation from simpler components of purine bases, one of the two classes of nitrogen-containing ring compounds found in DNA and RNA, which include adenine and guanine.
GO:0015172	acidic amino acid transporter activity molecular_function Enables the directed movement of acidic amino acids, amino acids with a pH below 7, into, out of, within or between cells.
GO:0015175	neutral amino acid transporter activity molecular_function Enables the directed movement of neutral amino acids, amino acids with no net charge, into, out of,

	within or between cells.
GO:0015302	oxidative phosphorylation uncoupler activity molecular_function
GO:0015367	oxoglutarate:malate antiporter activity molecular_function Catalysis of the reaction: oxoglutarate(out) + malate(in) = oxoglutarate(in) + malate(out).
GO:0015399	primary active transporter activity molecular_function Catalysis of transport of a solute against a concentration gradient using a primary energy source. Primary energy sources known to be coupled to transport are chemical, electrical and solar sources.
GO:0016071	mRNA metabolism biological_process The chemical reactions involving mRNA, messenger RNA, which is responsible for carrying the coded genetic 'message', transcribed from DNA, to sites of protein assembly at the ribosomes.
GO:0016117	carotenoid biosynthesis biological_process The formation from simpler components of carotenoids, tetraterpenoid compounds in which two units of 4 isoprenoid residues joined head-to-tail are themselves joined tail-to-tail.
GO:0016166	phytoene dehydrogenase activity molecular_function Catalysis of the dehydrogenation of phytoene to produce a carotenoid intermediate such as phytofluene.
GO:0016321	female meiosis chromosome segregation biological_process The process by which genetic material, in the form of chromosomes, is physically separated into two or more sets during meiosis in a female.
GO:0016804	prolyl aminopeptidase activity molecular_function Catalysis of the release of a N-terminal proline from a peptide.
GO:0017068	glutamyl-tRNA(Gln) amidotransferase activity molecular_function
GO:0017119	Golgi transport complex cellular_component A complex of proteins that, in vitro, stimulates intra-Golgi transport; a 13S complex, about 800 kDa in

	size and consists of at least five polypeptides. In yeast, this complex is called the Sec34/35 complex and is composed of eight subunits (Sec34p, Sec35p, Dor1p, Cod1p, Cod2p, Cod3p, Cod4p, and Cod5p).
GO:0019357	nicotinate nucleotide biosynthesis biological_process The formation from simpler components of nicotinamide nucleotides, any nucleotide that contains combined nicotinate (pyridine 3-carboxylic acid).
GO:0019358	nicotinate nucleotide salvage biological_process The generation of nicotinate nucleotide without de novo synthesis.
GO:0019379	sulfate assimilation, phosphoadenylyl sulfate reduction by phosphoadenylyl-sulfate reductase (thioredoxin) biological_process The pathway by which inorganic sulfate is processed and incorporated into sulfated compounds, where the phosphoadenylyl sulfate reduction step is catalyzed by the enzyme phosphoadenylyl-sulfate reductase (thioredoxin) (EC:1.8.4.8).
GO:0030101	natural killer cell activation biological_process The change in morphology and behavior of a natural killer cell in response to a cytokine, chemokine, cellular ligand, pathogen, or soluble factor.
GO:0030431	sleep biological_process The processes by which an organism enters and maintains a periodic, readily reversible state of reduced awareness and metabolic activity. Usually accompanied by physical relaxation, the onset of sleep in humans and other mammals is marked by a change in the electrical activity of the brain.
GO:0030794	(S)-coclaurine-N-methyltransferase activity molecular_function Catalysis of the reaction: S-adenosyl-L-methionine + (S)-coclaurine = S-adenosyl-L-homocysteine + (S)-N-methylcoclaurine.
GO:0040025	vulval development (sensu Nematoda) biological_process Processes aimed at the

	progression of the vulva over time, from its formation to the mature structure. Vulval development begins during the larval stages to give rise to a fully formed vulva in the adult. As in, but not restricted to, the roundworms (Nematoda, ncbi_taxonomy_id:6231).
GO:0046592	polyamine oxidase activity molecular_function Catalysis of the reaction: N1-acetylspermine + O2 + H2O = N1-acetylspermidine + 3-aminopropanal + H2O2.
GO:0048148	behavioral response to cocaine biological_process A change in the behavior of an organism as a result of exposure to cocaine.

List of significantly over/under-represented GOID's and corresponding text for genes affected by treatment when exposed to 1 hr of UV.

Table 10. Significantly Over/Under-Represented GOID's in 4 hr Significant Gene List for a 1% False Discovery Rate

GOID	Text
GO:0000059	protein-nucleus import, docking
GO:0000184	mRNA catabolism, nonsense-mediated decay biological_process The nonsense-mediated mRNA decay pathway degrades mRNAs transcribed from genes in which an amino-acid codon has changed to a nonsense codon; this prevents the translation of such mRNAs into truncated, and potentially harmful, proteins.
GO:0000290	deadenylylation-dependent decapping biological_process Cleavage of the 5'-cap of an mRNA triggered by shortening of the poly(A) tail to below a minimum functional length.
GO:0000311	plastid large ribosomal subunit cellular_component The larger of the two subunits of a plastid ribosome. Two sites on the ribosomal large subunit are involved in translation: the aminoacyl site (A site) and peptidyl site (P site).
GO:0000932	cytoplasmic mRNA processing body cellular_component A focus in the cytoplasm where mRNAs may become inactivated by decapping or some other mechanism. mRNA processing and binding proteins are localized to these foci.
GO:0001501	skeletal development biological_process Processes aimed at the progression of the skeleton over time, from its formation to the mature structure. The skeleton is the bony framework of the body in vertebrates (endoskeleton) or the hard outer envelope of insects (exoskeleton or dermoskeleton).
GO:0003735	structural constituent of ribosome molecular_function The action of a molecule that contributes to the structural integrity of the ribosome.
GO:0003862	3-isopropylmalate dehydrogenase activity molecular_function Catalysis of the reaction: 3-carboxy-2-hydroxy-4-methylpentanoate + NAD+ = 3-carboxy-4-methyl-2-oxopentanoate + NADH + H+.
GO:0003917	DNA topoisomerase type I activity

	molecular_function Catalysis of a DNA topological transformation by transiently cleaving one DNA strand at a time to allow passage of another strand; changes the linking number by +1 per catalytic cycle.
GO:0003929	RAN small monomeric GTPase activity molecular_function OBSOLETE. Catalysis of the reaction: GTP + H2O = GDP + phosphate.
GO:0003975	UDP-N-acetylglucosamine-dolichyl-phosphate N-acetylglucosaminephosphotransferase activity molecular_function Catalysis of the reaction: UDP-N-acetyl-D-glucosamine + dolichyl phosphate = UMP + N-acetyl-D-glucosaminyl-diphosphodolichol.
GO:0004096	catalase activity molecular_function Catalysis of the reaction: 2 H2O2 = O2 + 2 H2O.
GO:0004174	electron-transferring-flavoprotein dehydrogenase activity molecular_function Catalysis of the reaction: reduced ETF + ubiquinone = ETF + ubiquinol.
GO:0004396	hexokinase activity molecular_function Catalysis of the reaction: ATP + D-hexose = ADP + D-hexose 6-phosphate.
GO:0004450	isocitrate dehydrogenase (NADP+) activity molecular_function Catalysis of the reaction: isocitrate + NADP+ = 2-oxoglutarate + CO2 + NADPH + H+.
GO:0004591	oxoglutarate dehydrogenase (succinyl-transferring) activity molecular_function Catalysis of the reaction: 2-oxoglutarate + lipoamide = S-succinyldihydrolipoamide + CO2.
GO:0004681	oxoglutarate dehydrogenase (succinyl-transferring) activity molecular_function Catalysis of the reaction: 2-oxoglutarate + lipoamide = S-succinyldihydrolipoamide + CO2.
GO:0004815	aspartate-tRNA ligase activity molecular_function Catalysis of the reaction: ATP + L-aspartate + tRNA(Asp) = AMP + diphosphate + L-aspartyl-tRNA(Asp).
GO:0004872	receptor activity molecular_function

	Combining with an extracellular or intracellular messenger to initiate a change in cell activity.
GO:0005578	extracellular matrix (sensu Metazoa) cellular_component A layer consisting mainly of proteins (especially collagen) and glycosaminoglycans (mostly as proteoglycans) that forms a sheet underlying cells such as endothelial and epithelial cells. The proteins are secreted by cells in the vicinity. As in, but not restricted to, the multicellular animals (Metazoa, ncbi_taxonomy_id:33208).
GO:0005643	nuclear pore cellular_component Any of the numerous similar discrete openings in the nuclear envelope of a eukaryotic cell, where the inner and outer nuclear membranes are joined.
GO:0005840	ribosome cellular_component An intracellular organelle, about 200 A in diameter, consisting of RNA and protein. It is the site of protein biosynthesis resulting from translation of messenger RNA (mRNA). It consists of two subunits, one large and one small, each containing only protein and RNA. Both the ribosome and its subunits are characterized by their sedimentation coefficients, expressed in Svedberg units (symbol: S). Hence, the prokaryotic ribosome (70S) comprises a large (50S) subunit and a small (30S) subunit, while the eukaryotic ribosome (80S) comprises a large (60S) subunit and a small (40S) subunit. Two sites on the ribosomal large subunit are involved in translation, namely the aminoacyl site (A site) and peptidyl site (P site). Ribosomes from prokaryotes, eukaryotes, mitochondria, and chloroplasts have characteristically distinct ribosomal proteins.
GO:0005843	cytosolic small ribosomal subunit (sensu Eukaryota) cellular_component The small subunit of a eukaryotic cytosolic ribosome; has a sedimentation coefficient of 40S. As in, but not restricted to, the eukaryotes

	(Eukaryota, ncbi_taxonomy_id:2759).
GO:0006092	main pathways of carbohydrate metabolism biological_process
GO:0006323	DNA packaging biological_process Any process by which DNA and associated proteins are formed into a compact, orderly structure.
GO:0006405	RNA-nucleus export biological_process The directed movement of RNA from the nucleus to the cytoplasm.
GO:0006422	aspartyl-tRNA aminoacylation biological_process The process of coupling aspartate to aspartyl-tRNA, catalyzed by aspartyl-tRNA synthetase. In tRNA aminoacylation, the amino acid is first activated by linkage to AMP and then transferred to either the 2'- or the 3'-hydroxyl group of the 3'-adenosine residue of the tRNA.
GO:0006611	protein-nucleus export biological_process The directed movement of a protein from the nucleus into the cytoplasm.
GO:0006740	NADPH regeneration biological_process
GO:0006749	glutathione metabolism biological_process The chemical reactions involving glutathione, the tripeptide glutamylcysteinylglycine, which acts as a coenzyme for some enzymes and as an antioxidant in the protection of sulfhydryl groups in enzymes and other proteins; it has a specific role in the reduction of hydrogen peroxide (H2O2) and oxidized ascorbate, and it participates in the gamma-glutamyl cycle.
GO:0006997	nuclear organization and biogenesis biological_process The assembly and arrangement of the nucleus.
GO:0007052	mitotic spindle organization and biogenesis biological_process The formation and maintenance of the microtubule spindle during a mitotic cell cycle.
GO:0007166	cell surface receptor linked signal transduction biological_process Any series of molecular signals initiated by the binding of an extracellular ligand to a receptor on the surface of the target cell.

GO:0007224	smoothened signaling pathway biological_process The series of molecular signals generated as a consequence of activation of the transmembrane protein Smoothened.
GO:0008054	cyclin catabolism biological_process The breakdown into simpler components of cyclins, proteins whose levels in a cell varies markedly during the cell cycle, and which play key roles in regulating cell cycle phase transitions.
GO:0008122	amine oxidase activity molecular_function Catalysis of the reaction: R-CH2-NH2 + H2O + O2 = R-CHO + NH3 + H2O2.
GO:0008151	cellular physiological process biological_process The processes pertinent to the integrated function of a cell.
GO:0008430	selenium binding molecular_function Interacting selectively with selenium (Se).
GO:0015172	acidic amino acid transporter activity molecular_function Enables the directed movement of acidic amino acids, amino acids with a pH below 7, into, out of, within or between cells.
GO:0015175	neutral amino acid transporter activity molecular_function Enables the directed movement of neutral amino acids, amino acids with no net charge, into, out of, within or between cells.
GO:0015399	primary active transporter activity molecular_function Catalysis of transport of a solute against a concentration gradient using a primary energy source. Primary energy sources known to be coupled to transport are chemical, electrical and solar sources.
GO:0016071	mRNA metabolism biological_process The chemical reactions involving mRNA, messenger RNA, which is responsible for carrying the coded genetic 'message', transcribed from DNA, to sites of protein assembly at the ribosomes.
GO:0016117	carotenoid biosynthesis biological_process The formation from simpler components of carotenoids, tetraterpenoid compounds in which two units of 4 isoprenoid residues

	joined head-to-tail are themselves joined tail-to-tail.
GO:0016120	carotene biosynthesis biological_process The formation from simpler components of carotenes, hydrocarbon carotenoids.
GO:0016166	phytoene dehydrogenase activity molecular_function Catalysis of the dehydrogenation of phytoene to produce a carotenoid intermediate such as phytofluene.
GO:0016321	female meiosis chromosome segregation biological_process The process by which genetic material, in the form of chromosomes, is physically separated into two or more sets during meiosis in a female.
GO:0016804	prolyl aminopeptidase activity molecular_function Catalysis of the release of a N-terminal proline from a peptide.
GO:0019649	formaldehyde assimilation biological_process
GO:0030162	regulation of proteolysis and peptidolysis biological_process Any process that modulates the frequency, rate or extent of protein degradation by peptide bond hydrolysis.
GO:0042744	hydrogen peroxide catabolism biological_process The breakdown into simpler components of hydrogen peroxide (H2O2).
GO:0045475	locomotor rhythm biological_process The rhythm of the locomotor activity of an organism during its 24 hr activity cycle.

List of significantly over/under-represented GOID's and corresponding text for genes affected by treatment when exposed to 4 hr of UV.

Table 11. Significantly Over/Under-Represented GOID's in Treatment Profile 2 (Decreasing) for a 1% False Discovery Rate

GOID	Text
GO:0000166	nucleotide binding molecular_function Interacting selectively with a nucleotide, any compound consisting of a nucleoside that is esterified with (ortho)phosphate or an oligophosphate at any hydroxyl group on the glycose moiety.
GO:0000287	magnesium ion binding molecular_function Interacting selectively with magnesium (Mg) ions.
GO:0000795	synaptonemal complex cellular_component A proteinaceous scaffold found between homologous chromosomes during meiosis.
GO:0000812	SWR1 complex cellular_component A multisubunit protein complex that is involved in chromatin remodeling. It is required for the incorporation of the histone variant H2AZ into chromatin. In S. cerevisiae, the complex contains Swr1p, a Swi2/Snf2-related ATPase, and 12 additional subunits.
GO:0003678	DNA helicase activity molecular_function Catalysis of the hydrolysis of ATP to unwind the DNA helix at the replication fork, allowing the resulting single strands to be copied.
GO:0003763	chaperonin ATPase activity molecular_function OBSOLETE. Catalysis of the reaction: ATP + H2O = ADP + phosphate. The hydrolysis of ATP involved in maintaining an unfolded polypeptide structure before folding or to entry into mitochondria and chloroplasts.
GO:0003773	heat shock protein activity molecular_function OBSOLETE. Any of a group of specific proteins that are synthesized by both prokaryotic and eukaryotic cells after they have been exposed to a temperature that is higher than normal. Other stresses, e.g. free radical damage, have a similar effect. Many members of the hsp family are not induced but are present in all cells. They are characterized by their role as molecular

	chaperones.
GO:0003787	actin depolymerizing activity molecular_function OBSOLETE (was not defined before being made obsolete).
GO:0004478	methionine adenosyltransferase activity molecular_function Catalysis of the reaction: ATP + L-methionine + H2O = phosphate + diphosphate + S-adenosyl-L-methionine.
GO:0004506	squalene monooxygenase activity molecular_function Catalysis of the formation of a 2,3-epoxide in squalene.
GO:0004685	calcium- and calmodulin-dependent protein kinase activity molecular_function Catalysis of the reaction: ATP + protein = ADP + O-phosphoprotein.
GO:0005507	copper ion binding molecular_function Interacting selectively with copper (Cu) ions.
GO:0006122	mitochondrial electron transport, ubiquinol to cytochrome c biological_process The transfer of electrons from ubiquinol to cytochrome c that occurs during oxidative phosphorylation, mediated by the multisubunit enzyme known as complex III.
GO:0006310	DNA recombination biological_process The processes by which a new genotype is formed by reassortment of genes resulting in gene combinations different from those that were present in the parents. In eukaryotes genetic recombination can occur by chromosome assortment, intrachromosomal recombination, or nonreciprocal interchromosomal recombination. Intrachromosomal recombination occurs by crossing over. In bacteria it may occur by genetic transformation, conjugation, transduction, or F-duction.
GO:0006338	chromatin remodeling biological_process Dynamic structural changes to eukaryotic chromatin occurring throughout the cell division cycle. These changes range from the local changes necessary for transcriptional regulation to global changes

	necessary for chromosome segregation.
GO:0006520	amino acid metabolism biological_process The chemical reactions involving amino acids, organic acids containing one or more amino substituents.
GO:0006555	methionine metabolism biological_process The chemical reactions involving methionine (2-amino-4-(methylthio)butanoic acid), a sulfur-containing, essential amino acid found in peptide linkage in proteins.
GO:0006556	S-adenosylmethionine biosynthesis biological_process The formation from simpler components of S-adenosylmethionine, S-(5'-adenosyl)-L-methionine, an important intermediate in one-carbon metabolism.
GO:0006606	protein-nucleus import biological_process The directed movement of a protein from the cytoplasm to the nucleus.
GO:0006730	one-carbon compound metabolism biological_process The chemical reactions involving compounds containing a single carbon atom.
GO:0006869	lipid transport biological_process The directed movement of lipids into, out of, within or between cells. Lipids are compounds soluble in an organic solvent but not, or sparingly, in an aqueous solvent.
GO:0007098	centrosome cycle biological_process The process of centrosome duplication and separation. The centrosome cycle can operate with a considerable degree of independence from other processes of the cell cycle.
GO:0007143	female meiosis biological_process Meiosis in the female germline.
GO:0007283	spermatogenesis biological_process The process of formation of spermatozoa, including spermatocytogenesis and spermiogenesis.
GO:0007352	zygotic determination of dorsal/ventral axis biological_process The specification of the dorsal/ventral axis of the embryo, through the products of genes expressed in the zygote.

GO:0007465	R7 cell fate commitment biological_process The process during which the R7 photoreceptor commits to its cell fate. The R7 receptor contributes the central part of the rhabdomere in the apical parts of the ommatidium.
GO:0008077	Hsp70/Hsp90 organizing protein activity molecular_function OBSOLETE (was not defined before being made obsolete).
GO:0008080	N-acetyltransferase activity molecular_function Catalysis of the transfer of an acetyl group to a nitrogen atom on the acceptor molecule.
GO:0008121	ubiquinol-cytochrome-c reductase activity molecular_function Catalysis of the reaction: CoQH2 + 2 ferricytochrome c = CoQ + 2 ferrocytochrome c.
GO:0008289	lipid binding molecular_function Interacting selectively with a lipid.
GO:0008293	torso signaling pathway biological_process The series of molecular signals generated as a consequence of the torso transmembrane receptor tyrosine kinase binding to its physiological ligand.
GO:0008595	determination of anterior/posterior axis, embryo biological_process The specification of the anterior/posterior axis of the embryo by the products of genes expressed maternally and genes expressed in the zygote.
GO:0008638	protein tagging activity molecular_function OBSOLETE. Covalent addition of a specific tagging molecule to a protein, targeting the tagged protein for some fate e.g. degradation.
GO:0009269	response to dessication biological_process A change in state or activity of an organism (in terms of movement, secretion, gene expression, enzyme production, etc.) as a result of extreme dryness resulting from the prolonged deprivation of water.
GO:0009408	response to heat biological_process A change in state or activity of an organism (in terms of movement, secretion, gene expression, enzyme production, etc.) in response to temperatures above the optimal

	temperature for that organism.
GO:0009555	male gametophyte development biological_process Processes aimed at the progression of the male gametophyte over time, from its formation to the mature structure. This begins with the haploid microspores and ends with the mature microgametophytes containing the gametes (pollen).
GO:0009626	hypersensitive response biological_process The rapid death of plant cells in response to invasion by a pathogen.
GO:0009644	response to high light intensity biological_process
GO:0009793	embryonic development (sensu Magnoliophyta) biological_process Processes aimed at the progression of the embryo over time, from zygote formation to the end of seed dormancy. As in, but not restricted to, the flowering plants (Magnoliophyta, ncbi_taxonomy_id:3398).
GO:0009846	pollen germination biological_process The physiological and developmental changes that occur in a heterosporous plant pollen grain, beginning with hydration and terminating with the emergence of the pollen tube through the aperture.
GO:0009860	pollen tube growth biological_process Growth of pollen via tip extension of the intine wall.
GO:0015250	water channel activity molecular_function
GO:0016126	sterol biosynthesis biological_process The formation from simpler components of sterols, steroids with one or more hydroxyl groups and a hydrocarbon side-chain in the molecule.
GO:0016226	iron-sulfur cluster assembly biological_process The incorporation of iron and exogenous sulfur into a metallo-sulfur cluster.
GO:0016585	chromatin remodeling complex cellular_component Any complex that mediates dynamic changes in eukaryotic chromatin.
GO:0019592	mannitol catabolism biological_process The breakdown into simpler components of

	mannitol, the alditol derived from D-mannose by reduction of the aldehyde group.
GO:0019949	SUMO conjugating enzyme activity molecular_function Catalysis of the covalent attachment of the ubiquitin-like protein SUMO to other proteins.
GO:0030235	nitric-oxide synthase regulator activity molecular_function Modulates the activity of nitric oxide synthase.
GO:0030911	TPR domain binding molecular_function Interacting selectively with a tetratricopeptide repeat (TPR) domain of a protein, the consensus sequence of which is defined by a pattern of small and large hydrophobic amino acids and a structure composed of helices.
GO:0042026	protein refolding biological_process The process carried out by a cell that restores the biological activity of an unfolded or misfolded protein, using helper proteins such as chaperones.
GO:0042538	hyperosmotic salinity response biological_process A change in state or activity of an organism or cell (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of the perception of an increase in the concentration of salt (particularly but not exclusively sodium and chloride ions) in the environment.
GO:0042803	protein homodimerization activity molecular_function Interacting selectively with an identical protein to form a homodimer.
GO:0045187	regulation of circadian sleep/wake cycle, sleep biological_process Any process that modulates the frequency, rate or extent of sleep; a readily reversible state of reduced awareness and metabolic activity that occurs periodically in many animals.
GO:0045429	positive regulation of nitric oxide biosynthesis biological_process Any process that activates or increases the rate of the formation from simpler components of nitric oxide.

GO:0045941	positive regulation of transcription biological_process Any process that activates or increases the rate of transcription.
GO:0046029	mannitol dehydrogenase activity molecular_function Catalysis of the reaction: D-mannitol + NAD+ = D-mannose + NADH + H+.
GO:0047793	cycloeucalenol cycloisomerase activity molecular_function Catalysis of the reaction: cycloeucalenol = obtusifoliol.

List of significantly over/under-represented GOID's and corresponding text for genes affected by treatment and present in treatment profile 2, a decreasing pattern of expression across time.

Table 12. Significantly Over/Under-Represented GOID's in Treatment Profile 4 (Umbrella profile, downturn at 4 hr of UV) for a 1% False Discovery Rate

GOID	Text
GO:0000234	phosphoethanolamine N-methyltransferase activity molecular_function Catalysis of the reaction: S-adenosyl-L-methionine + ethanolamine phosphate = S-adenosyl-L-homocysteine + N-methylethanolamine phosphate.
GO:0001501	skeletal development biological_process Processes aimed at the progression of the skeleton over time, from its formation to the mature structure. The skeleton is the bony framework of the body in vertebrates (endoskeleton) or the hard outer envelope of insects (exoskeleton or dermoskeleton).
GO:0003926	ARF small monomeric GTPase activity molecular_function OBSOLETE. Catalysis of the reaction: GTP + H2O = GDP + phosphate.
GO:0003962	cystathionine gamma-synthase activity molecular_function Catalysis of the reaction: O-succinyl-L-homoserine + L-cysteine = cystathionine + succinate.
GO:0004568	chitinase activity molecular_function Catalysis of the random hydrolysis of N-acetyl-beta-D-glucosaminide 1,4-beta-linkages in chitin and chitodextrins.
GO:0004601	peroxidase activity molecular_function Catalysis of the reaction: donor + H2O2 = oxidized donor + 2 H2O.
GO:0005554	molecular_function unknown molecular_function Used for the annotation of gene products whose function is not known or cannot be inferred.
GO:0005786	signal recognition particle (sensu Eukaryota) cellular_component A ribonucleoprotein particle of 325 kDa composed of a 7S (300 nucleotide) RNA molecule and a complex of six different polypeptides. This binds both to the N-terminal signal peptide for proteins destined for the endoplasmic reticulum as they emerge from the large ribosomal subunit and also to the ribosome. This binding arrests further translation thereby

	preventing the proteins from being released into the cytosol. The SRP-ribosome complex then diffuses to the endoplasmic reticulum where it is bound to the signal recognition particle receptor, which allows resumption of protein synthesis and facilitates the passage of the growing polypeptide chain through the translocon. Through a process involving GTP hydrolysis, the SRP-SRP receptor complex dissociates and SRP returns to the cytosol. Of the six polypeptides of SRP the 54 kDa subunit (SRP54) is the central player. It contains an N-terminal GTPase domain and a C-terminal domain that binds directly to the signal peptide and the SRP RNA. As in, but not restricted to, the eukaryotes (Eukaryota, ncbi_taxonomy_id:2759).
GO:0005792	microsome cellular_component Any of the small, heterogeneous, artifactual, vesicular particles, 50-150 nm in diameter, that are formed when some eukaryotic cells are homogenized and that sediment on centrifugation at 100000 g.
GO:0006605	protein targeting biological_process The process of targeting specific proteins to particular membrane-bound subcellular organelles. Usually requires an organelle specific protein sequence motif.
GO:0006800	oxygen and reactive oxygen species metabolism biological_process The chemical reactions involving dioxygen (O2), or any of the reactive oxygen species, e.g. superoxide anions (O2-), hydrogen peroxide (H2O2), and hydroxyl radicals (-OH).
GO:0006869	lipid transport biological_process The directed movement of lipids into, out of, within or between cells. Lipids are compounds soluble in an organic solvent but not, or sparingly, in an aqueous solvent.
GO:0006917	induction of apoptosis biological_process A process that directly activates any of the steps required for cell death by apoptosis.
GO:0007252	I-kappaB phosphorylation biological_process The process of

	introducing a phosphoric group into an I-kappaB protein.
GO:0007582	physiological process biological_process Those processes specifically pertinent to the functioning of integrated living units: cells, tissues, organs, and organisms.
GO:0008154	actin polymerization and/or depolymerization biological_process Assembly or disassembly of actin filaments by the addition or removal of actin monomers from a filament.
GO:0008189	apoptosis inhibitor activity molecular_function OBSOLETE. The function held by products which directly block any step in the process of apoptosis.
GO:0008289	lipid binding molecular_function Interacting selectively with a lipid.
GO:0008379	thioredoxin peroxidase activity molecular_function Catalysis of the reaction: reduced thioredoxin + H2O2 = oxidized thioredoxin + H2O.
GO:0008430	selenium binding molecular_function Interacting selectively with selenium (Se).
GO:0008548	signal-recognition-particle GTPase activity molecular_function OBSOLETE. Catalysis of the reaction: GTP + H2O = GDP + phosphate. Activity is associated with the signal-recognition particle, a protein and RNA-containing structure involved in endoplasmic reticulum-associated protein synthesis.
GO:0008785	alkyl hydroperoxide reductase activity molecular_function Catalysis of the reaction: octane hydroperoxide + NADH + H+ = H2O + NAD+ + 1-octanol.
GO:0008964	phosphoenolpyruvate carboxylase activity molecular_function Catalysis of the reaction: phosphate + oxaloacetate = H2O + phosphoenolpyruvate + CO2.
GO:0009086	methionine biosynthesis biological_process The formation from simpler components of methionine (2-amino-4-(methylthio)butanoic acid), a sulfur-containing, essential amino acid found in peptide linkage in proteins.
GO:0009507	chloroplast cellular_component A

	chlorophyll-containing plastid with thylakoids organized into grana and frets, or stroma thylakoids, and embedded in a stroma.
GO:0009534	thylakoid (sensu Viridiplantae) cellular_component Sac-like membranous structures (cisternae) in a chloroplast combined into stacks (grana) and present singly in the stroma (stroma thylakoids or frets) as interconnections between grana. As in, but not restricted to, green plants and algae (Viridiplantae, ncbi_taxonomy_id:33090).
GO:0009535	thylakoid membrane (sensu Viridiplantae) cellular_component The pigmented membrane of a thylakoid. As in, but not restricted to, green plants and algae (Viridiplantae, ncbi_taxonomy_id:33090).
GO:0009538	photosystem I reaction center cellular_component A photochemical system containing P700, the chlorophyll a dimer that functions as a primary electron donor. Functioning as a light-dependent plastocyanin-ferredoxin oxidoreductase, it transfers electrons from plastocyanin to ferredoxin.
GO:0009626	hypersensitive response biological_process The rapid death of plant cells in response to invasion by a pathogen.
GO:0009739	response to gibberellic acid stimulus biological_process A change in state or activity of a cell or an organism as a result of the perception of gibberellic acid.
GO:0010044	response to aluminum ion biological_process A change in state or activity of an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of exposure to aluminum ions.
GO:0015979	photosynthesis biological_process The synthesis by organisms of organic chemical compounds, especially carbohydrates, from carbon dioxide (CO2) using energy obtained from light rather than from the oxidation of chemical compounds.
GO:0016209	antioxidant activity molecular_function

	Inhibition of the reactions brought about by dioxygen (O2) or peroxides. Usually the antioxidant is effective because it can itself be more easily oxidized than the substance protected. The term is often applied to components that can trap free radicals, thereby breaking the chain reaction that normally leads to extensive biological damage.
GO:0016706	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors molecular_function
GO:0016853	isomerase activity molecular_function Catalysis of the geometric or structural changes within one molecule. Isomerase is the systematic name for any enzyme of EC class 5.
GO:0018401	peptidyl-proline hydroxylation to 4-hydroxy-L-proline biological_process The posttranslational modification of peptidyl-proline to form 4-hydroxy-L-proline; catalyzed by procollagen-proline,2-oxoglutarate-4-dioxygenase.
GO:0030076	light-harvesting complex cellular_component A protein-pigment complex that may be closely or peripherally associated to photosynthetic reaction centers that participate in harvesting and transferring radiant energy to the reaction center.
GO:0030101	natural killer cell activation biological_process The change in morphology and behavior of a natural killer cell in response to a cytokine, chemokine, cellular ligand, pathogen, or soluble factor.
GO:0030125	clathrin vesicle coat cellular_component A clathrin coat found on a vesicle.
GO:0030131	clathrin adaptor complex cellular_component A membrane adaptor complex that links clathrin to a membrane.
GO:0040018	positive regulation of body size biological_process Any process that activates or increases the rate of growth of

	an organism to reach its usual body size.

List of significantly over/under-represented GOID's and corresponding text for genes affected by treatment and present in treatment profile 4, a downturn at 4 hr of UV exposure.

Table 13. Significantly Over/Under-Represented GOID's in Treatment Profile 5 (Umbrella profile, downturn at 8 hr of UV) for a 1% False Discovery Rate

GOID	Text
GO:0000004	biological_process unknown biological_process Used for the annotation of gene products whose process is not known or cannot be inferred.
GO:0000234	phosphoethanolamine N-methyltransferase activity molecular_function Catalysis of the reaction: S-adenosyl-L-methionine + ethanolamine phosphate = S-adenosyl-L-homocysteine + N-methylethanolamine phosphate.
GO:0000325	vacuole (sensu Magnoliophyta) cellular_component Cells contain one or several vacuoles, that may have different functions from each other. Vacuoles have a diverse array of functions. They can act as a storage organelle for nutrients or waste products, as a degradative compartment, as a cost-effective way of increasing cell size, and as a homeostatic regulator controlling both turgor pressure and pH of the cytosol. As in, but not restricted to, the flowering plants (Magnoliophyta, ncbi_taxonomy_id:3398).
GO:0001666	response to hypoxia biological_process The alteration of cellular processes due to exposure to lowered oxygen tension.
GO:0003676	nucleic acid binding molecular_function Interacting selectively with any nucleic acid.
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds molecular_function
GO:0005576	extracellular region cellular_component The space external to the outermost structure of a cell. For cells without external protective or external encapsulating structures this refers to space outside of the plasma membrane. This term covers the host cell environment outside an intracellular parasite.
GO:0005985	sucrose metabolism biological_process The chemical reactions involving sucrose, the disaccharide fructofuranosyl-glucopyranoside.

GO:0005986	sucrose biosynthesis biological_process The formation from simpler components of sucrose, the disaccharide fructofuranosyl-glucopyranoside.
GO:0008194	UDP-glycosyltransferase activity molecular_function Catalysis of the transfer of a glycosyl group from a UDP-sugar to a small hydrophobic molecule.
GO:0008270	zinc ion binding molecular_function Interacting selectively with zinc (Zn) ions.
GO:0008299	isoprenoid biosynthesis biological_process The formation from simpler components of any isoprenoid compound, isoprene (2-methylbuta-1,3-diene) or compounds containing or derived from linked isoprene (3-methyl-2-butenylene) residues.
GO:0009058	biosynthesis biological_process The energy-requiring part of metabolism in which simpler substances are transformed into more complex ones, as in growth and other biosynthetic processes.
GO:0009409	response to cold biological_process A change in state or activity of an organism (in terms of movement, secretion, gene expression, enzyme production, etc.) in response to temperatures below the optimal temperature for that organism.
GO:0016157	sucrose synthase activity molecular_function Catalysis of the reaction: UDP-glucose + D-fructose = UDP + sucrose.
GO:0016757	transferase activity, transferring glycosyl groups molecular_function Catalysis of the transfer of a glycosyl group from one compound (donor) to another (acceptor).
GO:0045174	glutathione dehydrogenase (ascorbate) activity molecular_function Catalysis of the reaction: 2 glutathione + dehydroascorbate = glutathione disulfide + ascorbate.
GO:0046429	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase activity molecular_function Catalysis of the reaction: (E)-4-hydroxy-3-methylbut-2-en-1-yl diphosphate + H2O + protein-disulfide = 2-C-methyl-D-erythritol 2,4-

	cyclodiphosphate + protein-dithiol. Note that (E)-4-hydroxy-3-methylbut-2-en-1-yl diphosphate is an alternative way of naming 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate.

List of significantly over/under-represented GOID's and corresponding text for genes affected by treatment and present in treatment profile 5, a downturn at 8 hr of UV exposure.

Table 14. Significantly Over/Under-Represented GOID's in Treatment Profile 7 (Inverted Umbrella profile, upturn at 4 hr of UV) for a 1% False Discovery Rate

GOID	Text
GO:0000151	ubiquitin ligase complex cellular_component A protein complex that includes a ubiquitin-protein ligase (E3) and other proteins that may confer substrate specificity on the complex.
GO:0000165	MAPKKK cascade biological_process Cascade of at least three protein kinase activities culminating in the phosphorylation and activation of a MAP kinase. MAPKKK cascades lie downstream of numerous signaling pathways.
GO:0000211	protein degradation tagging activity molecular_function OBSOLETE. Covalent addition of polyubiquitin to another protein, targeting the tagged protein for destruction.
GO:0002009	morphogenesis of an epithelium biological_process Processes by which the anatomical structures of epithelia are generated and organized. Morphogenesis pertains to the creation of form. An epithelium is a sheet of closely packed cells arranged in one or more layers, that covers the outer surfaces of the body or lines any internal cavity or tube.
GO:0003867	4-aminobutyrate transaminase activity molecular_function Catalysis of the reaction: 4-aminobutanoate + 2-oxoglutarate = succinate semialdehyde + L-glutamate.
GO:0003989	acetyl-CoA carboxylase activity molecular_function Catalysis of the reaction: ATP + acetyl-CoA + HCO3- = ADP + phosphate + malonyl-CoA.
GO:0004015	adenosylmethionine-8-amino-7-oxononanoate transaminase activity molecular_function Catalysis of the reaction: S-adenosyl-L-methionine + 8-amino-7-oxononanoate = S-adenosyl-4-methylthio-2-oxobutanoate + 7,8-diaminononanoate.
GO:0004028	aldehyde dehydrogenase activity molecular_function Catalysis of the

	reaction: 3-chloroallyl aldehyde + H2O = 2 H+ + 2 e- + 3-chloroacrylic acid.
GO:0004029	aldehyde dehydrogenase (NAD) activity molecular_function Catalysis of the reaction: an aldehyde + NAD+ + H2O = an acid + NADH + H+.
GO:0004030	aldehyde dehydrogenase [NAD(P)+] activity molecular_function Catalysis of the reaction: an aldehyde + NAD(P)+ + H2O = an acid + NAD(P)H + H+.
GO:0004124	cysteine synthase activity molecular_function Catalysis of the reaction: O3-acetyl-L-serine + H2S = L-cysteine + acetate.
GO:0004185	serine carboxypeptidase activity molecular_function Catalysis of the hydrolysis of carboxyl terminal peptide linkages in oligopeptides or polypeptides; a serine residue is at the active center.
GO:0004462	lactoylglutathione lyase activity molecular_function Catalysis of the reaction: (R)-S-lactoylglutathione = glutathione + methylglyoxal.
GO:0004467	long-chain-fatty-acid-CoA ligase activity molecular_function Catalysis of the reaction: ATP + a long-chain carboxylic acid + CoA = AMP + diphosphate + an acyl-CoA.
GO:0004708	MAP kinase kinase activity molecular_function Catalysis of the concomitant phosphorylation of threonine (T) and tyrosine (Y) residues in a Thr-Glu-Tyr (TEY) thiolester sequence in MAP kinases. It is a dual-specificity mitogen-activated protein kinase kinase and requires activation by the serine/threonine kinase, MAP kinase kinase kinase.
GO:0004788	thiamin diphosphokinase activity molecular_function Catalysis of the reaction: ATP + thiamin = AMP + thiamin diphosphate.
GO:0004842	ubiquitin-protein ligase activity molecular_function Catalysis of the reaction: ATP + ubiquitin + protein lysine = AMP + diphosphate + protein N-ubiquityllysine. Catalyzes the mediation of

	substrate recognition in ubiquitin-mediated protein degradation; binds directly to the substrate and its cognate E2 (ubiquitin conjugating enzyme).
GO:0005048	signal sequence binding molecular_function Interacting selectively with a signal sequence, a specific peptide sequence found on protein precursors or mature proteins that dictates where the mature protein is localized.
GO:0005065	heterotrimeric G-protein molecular_function OBSOLETE (was not defined before being made obsolete).
GO:0005080	protein kinase C binding molecular_function Interacting selectively with protein kinase C.
GO:0005102	receptor binding molecular_function Interacting selectively with one or more specific sites on a receptor molecule, a macromolecule that undergoes combination with a hormone, neurotransmitter, drug or intracellular messenger to initiate a change in cell function.
GO:0005386	carrier activity molecular_function Catalysis of the transfer of a specific substance or related group of substances from one side of the membrane to the other.
GO:0005496	steroid binding molecular_function Interacting selectively with a steroid, any of a large group of substances that have in common a ring system based on 1,2-cyclopentanoperhydrophenanthrene.
GO:0005497	androgen binding molecular_function Interacting selectively with any androgen, male sex hormones.
GO:0005624	membrane fraction cellular_component That fraction of cells, prepared by disruptive biochemical methods, that includes the plasma and other membranes.
GO:0005730	nucleolus cellular_component A small, dense body one or more of which are present in the nucleus of eukaryotic cells. It is rich in RNA and protein, is not bounded by a limiting membrane, and is not seen during mitosis. Its prime function is the

	transcription of the nucleolar DNA into 45S ribosomal-precursor RNA, the processing of this RNA into 5.8S, 18S, and 28S components of ribosomal RNA, and the association of these components with 5S RNA and proteins synthesized outside the nucleolus. This association results in the formation of ribonucleoprotein precursors; these pass into the cytoplasm and mature into the 40S and 60S subunits of the ribosome.
GO:0005802	Golgi trans face cellular_component The concave side of the Golgi apparatus, from which secretory vesicles emerge.
GO:0005819	spindle cellular_component The array of microtubules and associated molecules that forms between opposite poles of a eukaryotic cell during mitosis or meiosis and serves to move the duplicated chromosomes apart.
GO:0006081	aldehyde metabolism biological_process The chemical reactions involving aldehydes, any organic compound with the formula R-CH=O.
GO:0006120	mitochondrial electron transport, NADH to ubiquinone biological_process The transfer of electrons from NADH to ubiquinone that occurs during oxidative phosphorylation, mediated by the multisubunit enzyme known as complex I.
GO:0006479	protein amino acid methylation biological_process The addition of a methyl group to a protein amino acid. A methyl group is derived from methane by the removal of a hydrogen atom.
GO:0006535	cysteine biosynthesis from serine biological_process The formation of cysteine from simpler components, including serine.
GO:0006568	tryptophan metabolism biological_process The chemical reactions involving tryptophan, the chiral amino acid 2-amino-3-(1H-indol-3-yl)propanoic acid.
GO:0006610	ribosomal protein-nucleus import biological_process The directed movement of a ribosomal protein from the cytoplasm

	into the nucleus, across the nuclear membrane.
GO:0006629	lipid metabolism biological_process The chemical reactions involving lipids, compounds soluble in an organic solvent but not, or sparingly, in an aqueous solvent. Includes fatty acids; neutral fats, other fatty-acid esters, and soaps; long-chain (fatty) alcohols and waxes; sphingoids and other long-chain bases; glycolipids, phospholipids and sphingolipids; and carotenes, polyprenols, sterols, terpenes and other isoprenoids.
GO:0006772	thiamin metabolism biological_process The chemical reactions involving thiamin (vitamin B1), a water soluble vitamin present in fresh vegetables and meats, especially liver.
GO:0006776	vitamin A metabolism biological_process The chemical reactions involving any of the vitamin A compounds, retinol, retinal (retinaldehyde) and retinoic acid, all of which are derivatives of beta-carotene.
GO:0007140	male meiosis biological_process Meiosis in the male germline.
GO:0007205	protein kinase C activation biological_process The series of molecular signals that results in the upregulation of protein kinase C activity in response to the signal.
GO:0008137	NADH dehydrogenase (ubiquinone) activity molecular_function Catalysis of the reaction: NADH + H+ + ubiquinone = NAD+ + ubiquinol.
GO:0008276	protein methyltransferase activity molecular_function Catalysis of the transfer of a methyl group (CH3-) to a protein.
GO:0008308	voltage-gated ion-selective channel activity molecular_function
GO:0008638	protein tagging activity molecular_function OBSOLETE. Covalent addition of a specific tagging molecule to a protein, targeting the tagged protein for some fate e.g. degradation.
GO:0008757	S-adenosylmethionine-dependent

	methyltransferase activity molecular_function Catalysis of the transfer of a methyl group from S-adenosyl-L-methionine to a substrate.
GO:0009088	threonine biosynthesis biological_process The formation from simpler components of threonine (2-amino-3-hydroxybutyric acid), a polar, uncharged, essential amino acid found in peptide linkage in proteins.
GO:0009102	biotin biosynthesis biological_process The formation from simpler components of biotin, cis-tetrahydro-2-oxothieno(3,4-d)imidazoline-4-valeric acid.
GO:0009524	phragmoplast cellular_component Fibrous structure (light microscope view) that arises between the daughter nuclei at telophase and within which the initial partition (cell plate), dividing the mother cell in two (cytokinesis), is formed. Appears at first as a spindle connected to the two nuclei, but later spreads laterally in the form of a ring. Consists of microtubules.
GO:0009628	response to abiotic stimulus biological_process A change in state or activity of an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of the perception of an abiotic (non-living) stimulus.
GO:0009705	vacuolar membrane (sensu Magnoliophyta) cellular_component The single cytoplasmic membrane enclosing the vacuole; as in, but not restricted to, the flowering plants (Magnoliophyta, ncbi_taxonomy_id:3398).
GO:0009865	pollen tube adhesion biological_process The process by which the pollen tube adheres to cells of the stigma and style.
GO:0009954	proximal/distal pattern formation biological_process The processes that result in proximal/distal pattern formation.
GO:0010167	response to nitrate biological_process A change in state or activity of an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of exposure to a nitrate.

GO:0018991	oviposition biological_process The deposition of eggs (either fertilized or not) upon a surface or into a medium such as water.
GO:0019005	SCF ubiquitin ligase complex cellular_component A protein complex usually consisting of three conserved polypeptides (named Skp1, Cdc53/Cul1, Rbx1/Hrt1 in yeast and mammals) and one variable polypeptide (F-box protein), which is involved in the targeting of proteins for proteasome degradation.
GO:0019343	cysteine biosynthesis via cystathione biological_process The formation from simpler components of cysteine, via the intermediate cystathione.
GO:0019344	cysteine biosynthesis biological_process The formation from simpler components of cysteine, 2-amino-3-mercaptopropanoic acid.
GO:0030162	regulation of proteolysis and peptidolysis biological_process Any process that modulates the frequency, rate or extent of protein degradation by peptide bond hydrolysis.
GO:0030326	embryonic limb morphogenesis biological_process Processes, occurring in the embryo, by which the anatomical structures of the limb are generated and organized. Morphogenesis pertains to the creation of form. A limb is an appendage of an animal used for locomotion or grasping.
GO:0040010	positive regulation of growth rate biological_process Any process that increases the rate of growth of all or part of an organism.
GO:0040015	negative regulation of body size biological_process Any process that stops, prevents or reduces the rate of growth of an organism to reach its usual body size.
GO:0040017	positive regulation of locomotion biological_process Any process that activates or increases the rate of locomotion of an organism.
GO:0042573	retinoic acid metabolism biological_process The chemical reactions

	involving retinoic acid, one of the three components that makes up vitamin A.
GO:0042574	retinal metabolism biological_process The chemical reactions involving retinal, a compound that plays an important role in the visual process in most vertebrates. In the retina, retinal combines with opsins to form visual pigments. Retinal is one of the forms of vitamin A.
GO:0042981	regulation of apoptosis biological_process Any process that modulates the occurrence or rate of cell death by apoptosis.
GO:0045116	protein neddylation biological_process Covalent attachment of the ubiqitin-like protein NEDD8 (RUB1) to another protein.

List of significantly over/under-represented GOID's and corresponding text for genes affected by treatment and present in treatment profile 7, upturn at 4 hr of UV exposure.

DISCUSSION

Timecourse analysis of microarray data has been used successfully to dissect stress sensitivity in human keratinocytes (Sesto *et al*., 2002). Using microarrays to examine gene expression in human skin cells in response to UV irradiation, these authors confirmed up-regulation of several genes and described processes previously unknown to be affected by UV irradiation. Using cDNA microarrays to dissect the underlying mechanisms of UV responses in maize, the present study also confirmed previously identified gene functions in ultraviolet radiation responses, and defined new functions of genes not previously known.

A previous preliminary study indicated the upregulation of six transcripts in response to UV irradiation (Casati & Walbot, 2004): AW065914, an omega-6 fatty acid desaturase; AW144935, a cytochrome b5; AI947856, a glutamine synthetase; AI948309 and AW231530, two ribosomal proteins; and AI861109, a putative protein. Following 2 hr of UV irradiation, it was documented that two transcripts AI948309 and AW134461 (clathrin assembly protein) were upregulated (Casati & Walbot, 2004). Significant gene lists from this study were compared to that of the preliminary study in maize and it was confirmed that AI948309 was also significantly differentially expressed after 2 hr of UV exposure; however AW134461 was differentially expressed after 4 hr of UV exposure, which was different from the previous result.

New expression patterns of the previously identified upregulated transcripts were identified in this timecourse study. All transcripts with the exception of AI947856 were significantly differentially expressed after 1 hr of UV exposure. After 2 hr of UV exposure, only three transcripts were differentially expressed; AI948309, AI861109, and

AW144935. I also observed that the same transcripts present after 1 hr of UV exposure were also significantly differentially expressed after 4 and 8 hr of UV exposure. And, after 12 hr of UV exposure, three previously identified transcripts; AW065914, AQ144935, and AW231530 were differentially expressed in my significant gene lists.

Another study identified genes previously unknown to be affected by UV in the inbred parental lines Mo17 and B73 using a split-plot microarray design (Blum *et al*., 2004). For maize inbred line B73, genes identified in this previous study were compared to genes found to be significantly differentially expressed after UV in the present study. After 1 hr of UV irradiation, the following previously identified upregulated genes were significantly differentially expressed; AW224979, DNA/RNA binding proteins; AW399958, stress-related; BE012167, DNA binding; and BE056094, stress-related. Previously identified down-regulated genes were also significantly differentially expressed after 1 hr of UV exposure; AW289151, photosynthesis; AW352521, translation; and AW438057, DNA binding. The upregulation of genes involved in DNA binding and stress related responses as early as after 1 hr of UV exposure clearly indicates that the plant responds early to UV stress by activating proteins that bind to possibly damaged DNA. Down-regulation of a gene involved in translation (AW352521) implies that after 1 hr of UV exposure, the rate of translation is decreased to possibly inhibit the synthesis of damaged proteins. AW231338, a gene involved in signal transduction was previously identified as down-regulated and was identified in my study to be significantly differentially expressed after 2 and 4 hr of UV exposure. Not until 8 hr of UV exposure is there the presences of genes involved in signal transduction that are upregulated. Genes involved in photosynthesis were upregulated (AW062052) and

down-regulated (BE025333) after 1 hr of UV exposure, after 8 hr of UV exposure (AW288857, BE025281) and after 12 hr of UV exposure (AW289151, BE025281). After 8 hr of UV exposure, transcription factors and genes involved with translation are upregulated. This implies that after 8 hr of UV exposure, the plants normal processes involving DNA and protein synthesis begin to occur again.

Overview of Coordinated Gene Regulation

To date, many studies exist that investigate the effects of UV irradiation on gene expression, but the examination of coordinated gene expression in response to UV irradiation has been examined only by clustering. ORIOGEN, the profile analysis program utilized in this study, permitted the correct identification of coordinately regulated genes.

Eight treatment profile gene lists were analyzed to determine whether there were any significantly over/under-represented gene ontologies. Four treatment profiles were found to have over/under-represented gene ontologies and they were as follows; Profile 2 (Decreasing), Profile 4 (Umbrella profile, downturn at 4 hr of UV), Profile 5 (Umbrella profile, downturn at 8 hr of UV) and Profile 7 (Inverted Umbrella profile, upturn at 4 hr of UV).

Genes selected in Treatment profile 2 initially exhibited a high level of gene expression that decreased over time. One might expect that these genes are responsible for processes involved in the perception of UV, stress response, signal transduction, and they decrease across time as a result of plant acclimation. In fact, genes with over/under-represented gene ontologies present in this profile dealt specifically with these processes. Two genes were identified in the decreasing profile with processes known to be regulated

by UV but never identified as coordinately regulated by UV; AW927556, a gene responsible for the positive regulation of transcription and DNA recombination, and AW171907, a gene responsible for the positive regulation of nitric oxide biosynthesis and protein refolding. In bean leaves, increased durations of UV elevated hydrogen peroxide concentrations that were reduced by the production of nitric oxide (Shi *et al*., 2005). The reduction of hydrogen peroxide by nitric oxide not only reduced the production of harmful ROS in the cell but also alleviated UV-B induced damage. It was also proposed that nitric oxide acts as a signaling molecule mediating abiotic and biotic stress responses in plants (Shi *et al*., 2005). I propose that nitric oxide is acting as a second messenger upon UV stress activating the production of other biological molecules that respond to UV irradiation. Another gene identified in the decreasing profile functioned in protein folding. Protein dysfunction is a process known to be affected by UV irradiation. Proteins are vital cellular components that must maintain functional status to ensure cell survival. Several classes of heat shock proteins or chaperones have been identified to be induced by UV exposure that function in the proper folding, transport, and assembly of proteins in the cell (Wang et al., 2004). In this case, protein refolding is proposed to be an essential process upon UV stress.

Transcription, the process by which a DNA sequence is made into mRNA is also a process known to be affected by UV irradiation (Gyula *et al*., 2003) and it was identified in the decreasing profile. This was accompanied by DNA recombination, the process by which a new genotype is formed by the reassortment of genes on a chromosome (Russell, 2003). Previous literature has indicated that DNA recombination frequencies have been found to significantly increase in response to UV in *Arabidopsis*

thaliana (Molinier *et al*., 2005). Here it was found that genes involved with DNA recombination have high levels of gene expression initially, but over time the gene expression decreases as a result of prolonged UV exposure.

Profile 4 (Umbrella Profile, downturn at Time 3) is designated for genes with an increase in expression that decreased after 4 hr of UV exposure. Three genes were of interest in this profile; BQ048292, induction of apoptosis; AI948353, protein targeting; and BQ048292, antioxidant activity. There are several reasons these genes were of interest in this study. It was shown in *Arabidopsis thaliana* that exposure to UV induced apoptotic-like changes in the nuclei (Danon & Gallois, 1998). Apoptosis, programmed cell death, is a cell's last resort when repair cannot be accomplished. It appears that genes involved in this process are highly active until 4 hr of UV, at which point they decrease in expression.

As mentioned above, genes involved in protein folding were found in the decreasing profile. Interestingly, AI948353, a gene involved in protein targeting was found in profile 4 (Umbrella profile, downturn at 4 hr of UV). It appears that after 4 hr of UV, the majority of misfolded or damaged proteins have either been repaired or targeted for degradation. UV irradiation is a process that is known to produce harmful oxidants in the cell that are shielded by the production of flavonoids and other antioxidant compounds. Because BQ048292 was present in this profile, it was concluded that after 4 hr of antioxidant activity in response to UV decreases.

Treatment profile 5 (Umbrella profile, downturn at 4) included genes with increasing gene expression until 8 hr of UV at which point gene expression decreased. Several genes with functions in specific enzymatic activities were present in this list, but

the vacuole was found in this profile only corresponding to gene AW399990. Vacuoles are large vesicles that are used to store nutrients, metabolites, and waste products (Russell, 2003). It was shown in green algae that UV irradiation induced the formation of vacuoles in cells (Meindl & Lutz, 1996). After 8 hr of UV exposure, it is evident that the gene involved in the formation of vacuoles peaked in expression possibly coping with the large amount of waste products produced from UV irradiation.

Profile 7 (Inverted Umbrella profile, upturn at time 3) is designated for genes with a decreasing pattern of expression that increased at 4 hr of UV. Profile 4 and Profile 7 are inversely related; Profile 4 increases until 4 hr at which point it decreases and profile 7 decreases until 4 hr at which point it increases. Therefore, it is expected that genes present in profile 4 are possible regulators of genes present in profile 7. AW461038, a gene involved with the response to abiotic stimuli was present in profile 7. This gene initially has a high level of expression and at 4 hr of UV exposure it drops to its lowest point.

Genes with Continuous Expression

Two genes were identified that were significantly differentially expressed across all five timepoints, AI674008 and BQ293537. Gene BQ293537 had no gene ontology information associated with it. However, there were three gene ontologies associated with gene AI674008. The functions associated with this gene were "intracellular protein transport, the directed movement of proteins in a cell, including the movement of proteins between specific compartments or structures within a cell, such as organelles of a eukaryotic cell," "protein carrier activity, the catalysis of the transfer of proteins from one side of the membrane to the other," "the membrane cellular component, double layer of

lipid molecules that encloses all cells, and, in eukaryotes , many organelles; may be a single or double lipid bilayer; also includes associated proteins." Based on the functions associated with this continuously expressed gene in addition to the overwhelming presence of genes involving protein functions in the profiles, it is evident that protein transport, integrity, and degradation is an essential process upon UV stress. It appears that the cell continuously transports proteins to the necessary locations in the cell to ensure the plants survival.

Overview of Gene Ontology Analysis

Significantly over/under-represented gene ontologies were analyzed to identify functions of genes affected by UV treatment. Overall, five significant gene lists were analyzed to identify over/under-represented gene ontologies, and two lists, treatment affected genes after 1 hr and 4 hr of UV, had them present. Following 1 hr of UV exposure, gene ontologies involving in purine metabolism, nuclear division, and tRNA binding were significantly over/under-represented. AI438424, a gene that involves the metabolism of oxygen and reactive oxygen species was over/under-represented after 1 hr of UV exposure. This gene was also found to be significantly differentially expressed after 1 hr of UV exposure exclusively. Reactive oxygen species are diatomic oxygen molecules with the potential to undergo reduction due to the presence of two uncoupled electrons (Bergamini *et al*., 2004). Exposure to irradiation has been found to adversely affect the cell by causing the production of these species of molecules (Carletti *et al*., 2003; Brosche & Strid, 2003; Jordan, 2002; Hollosy, 2002). The fact that this gene that functions in the metabolism of reactive oxygen species was present after only 1 hr of UV exposure exclusively indicates that a cellular response is quickly initiated upon

perception of ultraviolet radiation in the plant and the metabolism of reactive oxygen species is an essential process. The production of these molecules after 1 hr of UV exposure could possibly be the signal needed for the production of protective flavonoid pigments in the plant.

After 4 hr of UV exposure, genes involved with mRNA catabolism and nonsense-mediated decay, ribosomal subunits, and numerous instances of genes involved with specific enzymatic activities relating to amino acid biosynthesis were evident. There were quite a few gene ontologies present after 4 hr of UV exposure that dealt with ribosomes and mRNA as well. High levels of ultraviolet radiation have been found to damage ribosomes and decrease the amount of new protein production in Maize resulting in alterations in translation (Casati & Walbot, 2004). Because gene ontologies involved in these types of processes are significantly over/under-represented, it is believed that after 4 hr of UV exposure, the plant cells have perceived a signal and are responding to the signal by altering translation.

Overview of Gene Functional Analysis

Each treatment-significant gene list was analyzed by ALACK to determine if any significantly over-represented gene description terms were present. Enzymes associated with purine metabolism, the breakdown or biosynthesis of purines, were over-represented gene description terms present in the 1, 2, and 4 hr significant gene lists. This was expected because of the documented effect UV irradiation has on DNA (Stratmann, 2003; Frohnmeyer & Staiger, 2003; Stapleton, 1992; Brosche & Strid, 2003).

Comparison of profiles

Treatment profile lists were also analyzed to determine if significantly over-represented gene description terms were present. Of the eight profiles analyzed, treatment profile 7 (Upturn at 4 hr of UV exposure) was the only profile to have any over-represented gene description terms present. DNA, the most vulnerable molecule in an organism in response to ultraviolet radiation (Stratmann, 2003; Frohnmeyer & Staiger, 2003; Stapleton, 1992; Brosche & Strid, 2003), was over-represented.

CONCLUSIONS

The observed pattern of differential gene expression varied drastically across timepoints. After 1 hr of UV exposure differential gene expression increased slightly similar to that of differential gene expression after 2 hr, after 4 hr of UV exposure differential gene expression decreased dramatically, and after 8 and 12 hours, differential gene expression increased again and was relatively similar. The pattern of differential gene expression observed implied several things; the transcriptome responded rapidly to UV stress possibly due to the increased expression of genes involved specifically in UV stress response or 2nd effects, photosynthesis is significantly affected across time in response to UV, and translation and DNA synthesis is affected upon UV stress. Data mining and further analysis will have to be done to further analyze this tremendous amount of biological data presented in this study.

LITERATURE CITED

Ashburner M. & Lewis S. (2002) On ontologies for biologists: the Gene Ontology-untangling the web. *Novartis Found Symposium*. **247**: 66-90.

Bar-Joseph Z. (2004) Analyzing time series gene expression data. *Bioinformatics*. **20**: 2493-2503.

Bergamini C., Gambetti S., Dondi A. & Cervellati C. (2004) Oxygen, reactive oxygen species and tissue damage. *Current Pharmaceutical Design*. **10**: 1611-1626.

Beissbarth T. & Speed TP. (2004) GOstat: Find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics*. **20** (9): 1464-1465.

Blum J., Casati P., Walbot V. & Stapleton A.E. (2004) Split-plot microarray design allows sensitive detection of expression differences after ultraviolet radiation in the inbred parental lines of a key Maize mapping population. *Plant, Cell and Environment*. **27**: 1374-1386.

Brosche M., Schuler M.A., Kalbina I., Connor L. & Strid A. (2002) Gene regulation by low level UV-B irradiation: identification by DNA array analysis. *Photochemical and Photobiological Sciences*. **1**: 656-664.

Brosche M. & Strid A. (2003) Molecular events following perception of ultraviolet-B radiation by plants. *Physiologia Plantarum*. **117**: 1-10.

Carletti P., Masi A., Wonisch A., Grill D., Tausz M. & Ferretti M. (2003) Changes in antioxidant and pigment pool dimensions in UV-B irradiated Maize seedlings. *Environmental and Experimental Botany*. **50**: 149-157.

Cartwright H.N., Baucom C., Singh P., Smith K.L. & Stapleton A.E. (2001) Intraspecific comparisons reveal differences in the pattern of ultraviolet radiation responses in four Maize (Zea mays L.) varieties. *Journal of Photochemistry and Photobiology*. **62**: 88-96.

Casati P. & Walbot V. (2003) Gene Expression Profiling in Response to Ultraviolet Radiation in Maize Genotypes with Varying Flavonoid Content. *Plant Physiology*. **132**: 1739-1754.

Casati P. & Walbot V. (2004) Rapid transcriptome responses of Maize (Zea mays) to UV- B in irradiated and shielded tissues. *Genome Biology*. **5**: R16.

Casati P. & Weinberg C.R. (2004) Crosslinking of Ribosomal Proteins to RNA in Maize Ribosomes by UV-B and Its Effects on Translation. *Plant Physiology*. **136**: 3319-3332.

Danon A. & Gallois P. (1998) UV-C radiation induces apoptotic-like changes in Arabidopsis thaliana. *FEBS Letters*. **437**: 131-136.

Frohnmeyer H. & Staiger D. (2003) Ultraviolet-B Radiation-Mediated Responses in Plants. Balancing Damage and Protection. *Plant Physiology*. **133**: 1420-1428.

Gentleman R.C., Carey V.J., Bates D.M., Bolstad B., Dettling M., Dudoit S., Ellis B., Gautier L., Ge Y., Gentry J., Hornik K., Hothorn T., Huber W., Iacus S., Irizarry R., Leisch F., Li C., Maechler M., Rossini A.J., Sawitzki G., Smith C., Smyth G., Tierney L., Yang J.Y. & Zhang J. (2004) Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biology*. **5**: R80.

Gyula P., Schafer E. & Nagy F. (2003) Light perception and signaling in higher plants. *Current Opinion in Plant Biology*. **6**: 446-452.

Hollosy F. (2002) Effects of ultraviolet radiation on plant cells. *Micron*. **33**: 179-197.

Huber W. (2004) Robust calibration and variance stabilization with VSN.

Jordan B.R. (2002) Molecular response of plant cells to UV-B stress. *Functional Plant Biology*. **29**: 909-916.

Kim C.C. & Falkow S. (2003) Significance analysis of lexical bias in microarray data. *BMC Bioinformatics*. **4**: 12.

Lee M.-L.T., Sharopova N., Beavis W.D., Grant D., Katt M., Blair D. & Hallauer A. (2002) Expanding the genetic map of Maize with the intermated B73 x Mo17 (IBM) population. *Plant Molecular Biology*. **48**: 453-461.

Lumsden, P. (1997) *Plants and UV-B responses to environmental change*. Cambridge University Press, UK, 355 pp.

Meindl U. & Lutz C. (1996) Effects of UV irradiation on cell development and ultrastructure of the green algae Micrasterias. *Journal of Photochemistry and Photobiology*. **36**: 285-292.

Molinier J., Oakeley E.J., Niederhauser O., Kovalchik I. & Hohn B. (2005) Dynamic response of plant genome to ultraviolet radiation and other genotoxic stresses. *Fundamental and Molecular Mechanisms of Mutagenesis*. **571**: 235-247.

Peddada S.D., Lobenhofer E.K., Li L., Afshari C.A., Weinberg C.R. & Umbach D.M. (2003) Gene selection and clustering for time-course and dose-response microarray experiments using order-restricted inference. *Bioinformatics*. **19**: 834-841.

Peng X., Wood C., Blalock E., Chen K., Landfield P. & Stromgerg A. (2003) Statistical implications of pooling RNA samples for microarray experiments. *BMC Bioinformatics*. **4**: 26.

Polidoros A.N. and Scandalios J.G. (1997) Response of the Maize catalases to light. *Free Radical Biology & Medicine*. **23** (3): 497-504.

R Development Core Team. (2004). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria ISBN 3-900051-00-3.

Russell, P. (2003) *Essential iGenetics*. Benjamin Cummings, USA, 614 pp.

Saeed A., Sharov V., White J., Li J., Liang W., Ghagabati N., Braisted J., Klapa M., Currier T., Thiagarajan M., Sturn A., Snuffin M., Rezantsev A., Popov D., Ryltsov A., Kostukovich E., Borisovsky I., Liu Z., Vinsavich A., Trush V. & Quackenbush J. (2003) TM4: a free, open-source system for microarray data management and analysis. *Biotechniques*. **34**: 374-378.

Schena M., Heller R.A., Theriault T.P., Konrad K., Lachenmeier E., Davis R.W. (1998) Microarrays: biotechnology's discovery platform for functional genomics. *Trends in Biotechnology*. **16**: 301-306.

Schomburg, I., Chang, A., Schomburg, D. (2002) BRENDA, enzyme data and metabolic information. *Nucleic Acids Research*. **30**: 47-9

Sinha R. & Hader D. (2002) UV-induced DNA damage and repair: a review. *Photochemical and Photobiological Sciences*. **4**: 225-236.

Stapleton, A. E. (1992). Ultraviolet Radiation and Plants: Burning Questions. *The Plant Cell*. **4**: 1353-1358.

Stapleton A.E, Walbot V. Flavonoids can protect Maize DNA from the induction of ultraviolet radiation damage. *Plant Physiology*. **105** (3) 881-889.

Storey, J. and R. Tibshirani (2003). Statistical significance for genomewide studies. *Proceedings of the National of Academy Sciences*. **100**(16): 9440-9445.

Stratmann, J. (2003). Ultraviolet-B radiation co-opts defense signaling pathways. *TRENDS in Plant Science*. **8**(11): 526-533.

Yang, Y. H., S. Dudoit, et al. (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Research*. **30**(4 e15): 1-10.

Zhu, T. (2003) Global analysis of gene expression using GeneChip microarrays. *Current Opinion in Plant Biology*. **6**: 418-425.

APPENDIX A

SAS Code

```
/* INPUTTING RAW DATA*/
data newa1tav;
infile 'C:\Carletha\Microarray Information\new1atav.csv' dlm = ",";
input Row Column Subrow SubCol RowwiSub ColwiSub IntAa    IntBa   Spotmeanratio
      Spotarea       Saturationfactor       Spotmedianratio       Spotmoderatio
      BackgroundA BackgroundB FlagA          FlagB  QCScoreA     QCScoreB
      TotalQCScore;
run;
data newb1tav;
infile 'C:\Carletha\Microarray Information\new1btav.csv' dlm = ",";
input Row Column Subrow SubCol RowwiSub ColwiSub IntAb IntBb Spotmeanratio
Spotarea       Saturationfactor Spotmedianratio Spotmoderatio BackgroundA
      BackgroundB  FlagA          FlagB  QCScoreA QCScoreB TotalQCScore;
run;
data newc1tav;
infile 'C:\Carletha\Microarray Information\new1ctav.csv' dlm = ",";
input Row Column Subrow SubCol RowwiSub ColwiSub IntAc    IntBc   Spotmeanratio
      Spotarea       Saturationfactor       Spotmedianratio       Spotmoderatio
      BackgroundA BackgroundB FlagA          FlagB  QCScoreA     QCScoreB
      TotalQCScore;
run;
data newd1tav;
infile 'C:\Carletha\Microarray Information\new1dtav.csv' dlm = ",";
input Row Column Subrow SubCol RowwiSub ColwiSub IntAd    IntBd Spotmeanratio
Spotarea Saturationfactor Spotmedianratio Spotmoderatio BackgroundA BackgroundB
FlagA   FlagB   QCScoreA QCScoreB TotalQCScore;
run;
data newa2tav;
infile 'C:\Carletha\Microarray Information\new2atav.csv' dlm = ",";
input Row Column Subrow SubCol RowwiSub ColwiSub IntAa    IntBa   Spotmeanratio
      Spotarea       Saturationfactor       Spotmedianratio       Spotmoderatio
      BackgroundA BackgroundB FlagA          FlagB  QCScoreA     QCScoreB
      TotalQCScore;
run;
data newb2tav;
infile 'C:\Carletha\Microarray Information\new2btav.csv' dlm = ",";
input Row Column Subrow SubCol RowwiSub ColwiSub IntAb    IntBb   Spotmeanratio
      Spotarea       Saturationfactor       Spotmedianratio       Spotmoderatio
      BackgroundA BackgroundB FlagA          FlagB  QCScoreA     QCScoreB
      TotalQCScore;
run;
data newc2tav;
infile 'C:\Carletha\Microarray Information\new2ctav.csv' dlm = ",";
```

```
input Row Column Subrow SubCol RowwiSub ColwiSub IntAc    IntBc   Spotmeanratio
      Spotarea      Saturationfactor      Spotmedianratio      Spotmoderatio
      BackgroundA BackgroundB FlagA         FlagB  QCScoreA    QCScoreB
      TotalQCScore;
run;
data newd2tav;
infile 'C:\Carletha\Microarray Information\new2dtav.csv' dlm = ",";
input Row Column Subrow SubCol RowwiSub ColwiSub IntAd    IntBd   Spotmeanratio
      Spotarea      Saturationfactor      Spotmedianratio      Spotmoderatio
      BackgroundA BackgroundB FlagA         FlagB  QCScoreA    QCScoreB
      TotalQCScore;
run;
data newa4tav;
infile 'C:\Carletha\Microarray Information\new4atav.csv' dlm = ",";
input Row Column Subrow SubCol RowwiSub ColwiSub IntAa    IntBa   Spotmeanratio
      Spotarea      Saturationfactor      Spotmedianratio      Spotmoderatio
      BackgroundA BackgroundB FlagA         FlagB  QCScoreA    QCScoreB
      TotalQCScore;
run;
data newb4tav;
infile 'C:\Carletha\Microarray Information\new4btav.csv' dlm = ",";
input Row Column Subrow SubCol RowwiSub ColwiSub IntAb    IntBb   Spotmeanratio
      Spotarea      Saturationfactor      Spotmedianratio      Spotmoderatio
      BackgroundA BackgroundB FlagA         FlagB  QCScoreA    QCScoreB
      TotalQCScore;
run;
data newc4tav;
infile 'C:\Carletha\Microarray Information\new4ctav.csv' dlm = ",";
input Row Column Subrow SubCol RowwiSub ColwiSub IntAc    IntBc   Spotmeanratio
      Spotarea      Saturationfactor      Spotmedianratio      Spotmoderatio
      BackgroundA BackgroundB FlagA         FlagB  QCScoreA    QCScoreB
      TotalQCScore;
run;
data newd4tav;
infile 'C:\Carletha\Microarray Information\new4dtav.csv' dlm = ",";
input Row Column Subrow SubCol RowwiSub ColwiSub IntAd    IntBd   Spotmeanratio
      Spotarea      Saturationfactor      Spotmedianratio      Spotmoderatio
      BackgroundA BackgroundB FlagA         FlagB  QCScoreA    QCScoreB
      TotalQCScore;
run;
data newa8tav;
infile 'C:\Carletha\Microarray Information\new8atav.csv' dlm = ",";
input Row Column Subrow SubCol RowwiSub ColwiSub IntAa    IntBa   Spotmeanratio
      Spotarea      Saturationfactor      Spotmedianratio      Spotmoderatio
      BackgroundA BackgroundB FlagA         FlagB  QCScoreA    QCScoreB
      TotalQCScore;
```

```
run;
data newb8tav;
infile 'C:\Carletha\Microarray Information\new8btav.csv' dlm = ",";
input Row Column Subrow SubCol RowwiSub ColwiSub IntAb IntBb Spotmeanratio
      Spotarea Saturationfactor Spotmedianratio Spotmoderatio
      BackgroundA BackgroundB FlagA FlagB QCScoreA QCScoreB
      TotalQCScore;
run;
data newc8tav;
infile 'C:\Carletha\Microarray Information\new8ctav.csv' dlm = ",";
input Row Column Subrow SubCol RowwiSub ColwiSub IntAc IntBc Spotmeanratio
      Spotarea Saturationfactor Spotmedianratio Spotmoderatio
      BackgroundA BackgroundB FlagA FlagB QCScoreA QCScoreB
      TotalQCScore;
run;
data newd8tav;
infile 'C:\Carletha\Microarray Information\new8dtav.csv' dlm = ",";
input Row Column Subrow SubCol RowwiSub ColwiSub IntAd IntBd Spotmeanratio
      Spotarea Saturationfactor Spotmedianratio Spotmoderatio
      BackgroundA BackgroundB FlagA FlagB QCScoreA QCScoreB
      TotalQCScore;
run;
data newa12tav;
infile 'C:\Carletha\Microarray Information\new12atav.csv' dlm = ",";
input Row Column Subrow SubCol RowwiSub ColwiSub IntAa IntBa Spotmeanratio
      Spotarea Saturationfactor Spotmedianratio Spotmoderatio
      BackgroundA BackgroundB FlagA FlagB QCScoreA QCScoreB
      TotalQCScore;
run;
data newb12tav;
infile 'C:\Carletha\Microarray Information\new12btav.csv' dlm = ",";
input Row Column Subrow SubCol RowwiSub ColwiSub IntAb IntBb Spotmeanratio
      Spotarea Saturationfactor Spotmedianratio Spotmoderatio
      BackgroundA BackgroundB FlagA FlagB QCScoreA QCScoreB
      TotalQCScore;
run;
data newc12tav;
infile 'C:\Carletha\Microarray Information\new12ctav.csv' dlm = ",";
input Row Column Subrow SubCol RowwiSub ColwiSub IntAc IntBc Spotmeanratio
      Spotarea Saturationfactor Spotmedianratio Spotmoderatio
      BackgroundA BackgroundB FlagA FlagB QCScoreA QCScoreB
      TotalQCScore;
run;
data newd12tav;
infile 'C:\Carletha\Microarray Information\new12dtav.csv' dlm = ",";
```

```
input Row Column Subrow SubCol RowwiSub ColwiSub IntAd    IntBd   Spotmeanratio
      Spotarea       Saturationfactor       Spotmedianratio       Spotmoderatio
      BackgroundA BackgroundB FlagA          FlagB  QCScoreA     QCScoreB
      TotalQCScore;
run;

/* REMOVE UNWANTED DATA WITH FLAGS */
data newa1tav;
set newa1tav;
if FlagA="Z" or FlagB="Z" then delete;
run;
data newb1tav;
set newb1tav;
if FlagA="Z" or FlagB="Z" then delete;
run;
data newc1tav;
set newc1tav;
if FlagA="Z" or FlagB="Z" then delete;
run;
data newd1tav;
set newd1tav;
if FlagA="Z" or FlagB="Z" then delete;
run;
data newa2tav;
set newa2tav;
if FlagA="Z" or FlagB="Z" then delete;
run;
data newb2tav;
set newb2tav;
if FlagA="Z" or FlagB="Z" then delete;
run;
data newc2tav;
set newc2tav;
if FlagA="Z" or FlagB="Z" then delete;
run;
data newd2tav;
set newd2tav;
if FlagA="Z" or FlagB="Z" then delete;
run;
data newa4tav;
set newa4tav;
if FlagA="Z" or FlagB="Z" then delete;
run;
data newb4tav;
set newb4tav;
if FlagA="Z" or FlagB="Z" then delete;
```

```
run;
data newc4tav;
set newc4tav;
if FlagA="Z" or FlagB="Z" then delete;
run;
data newd4tav;
set newd4tav;
if FlagA="Z" or FlagB="Z" then delete;
run;
data newa8tav;
set newa8tav;
if FlagA="Z" or FlagB="Z" then delete;
run;
data newb8tav;
set newb8tav;
if FlagA="Z" or FlagB="Z" then delete;
run;
data newc8tav;
set newc8tav;
if FlagA="Z" or FlagB="Z" then delete;
run;
data newd8tav;
set newd8tav;
if FlagA="Z" or FlagB="Z" then delete;
run;
data newa12tav;
set newa12tav;
if FlagA="Z" or FlagB="Z" then delete;
run;
data newb12tav;
set newb12tav;
if FlagA="Z" or FlagB="Z" then delete;
run;
data newc12tav;
set newc12tav;
if FlagA="Z" or FlagB="Z" then delete;
run;
data newd12tav;
set newd12tav;
if FlagA="Z" or FlagB="Z" then delete;
run;

/* INPUT UNIGENE FILES*/
data unilarge;
infile 'C:\Carletha\unigenelarge.csv' dsd truncover lrecl=10000;
```

```
input SubRow SubCol RowwiSub ColwiSub Gene $ Gi :  10. TUC_ID $ :  30.
Homo1_PID $ :  80. Homo1_Desc $ : 300. Homo1_Species $ :  80. Homo2_PID $ :  80.
Homo2_Desc $ :  80. Homo2_Species $ :  80. Homo3_PID $ :  80. Homo3_Desc $ :  80.
Homo3_Species $ :  80.;
run;
data unismall;
infile 'C:\Carletha\unigenesmall.csv' dsd truncover lrecl=10000;
input SubRow SubCol RowwiSub ColwiSub Gene $ Gi :  10. TUC_ID $ :  30.
Homo1_PID $ :  80. Homo1_Desc $ : 80. Homo1_Species $ :  80. Homo2_PID $ :  80.
Homo2_Desc $ :  80. Homo2_Species $ :  80. Homo3_PID $ :  80. Homo3_Desc $ :  80.
Homo3_Species $ :  80.;
run;

/* RENUMBERED UNIGENE SMALL*/
data unismall1;
set unismall;
if SubRow=1 then SubRow=9;
if SubRow=2 then SubRow=10;
if SubRow=3 then SubRow=11;
if SubRow=4 then SubRow=12;
run;
data unigene;
set unilarge unismall1;
length HOmo1_Desc $10000;
if gene='BLANK' then delete;
if gene='blank' then delete;
if gene='xxxx' then delete;
if gene='Blank' then delete;
run;
proc sort data=unigene;
by SubRow SubCol RowwiSub ColwiSub;
run;

/* SORTED TO MERGE DATA AND UNIGENE FILE*/
proc sort data=newa1tav;
by SubRow SubCol RowwiSub ColwiSub;
run;
proc sort data=newb1tav;
by SubRow SubCol RowwiSub ColwiSub;
run;
proc sort data=newc1tav;
by SubRow SubCol RowwiSub ColwiSub;
run;
proc sort data=newd1tav;
by SubRow SubCol RowwiSub ColwiSub;
run;
```

```
proc sort data=newa2tav;
by SubRow SubCol RowwiSub ColwiSub;
run;
proc sort data=newb2tav;
by SubRow SubCol RowwiSub ColwiSub;
run;
proc sort data=newc2tav;
by SubRow SubCol RowwiSub ColwiSub;
run;
proc sort data=newd2tav;
by SubRow SubCol RowwiSub ColwiSub;
run;
proc sort data=newa4tav;
by SubRow SubCol RowwiSub ColwiSub;
run;
proc sort data=newb4tav;
by SubRow SubCol RowwiSub ColwiSub;
run;
proc sort data=newc4tav;
by SubRow SubCol RowwiSub ColwiSub;
run;
proc sort data=newd4tav;
by SubRow SubCol RowwiSub ColwiSub;
run;
proc sort data=newa8tav;
by SubRow SubCol RowwiSub ColwiSub;
run;
proc sort data=newb8tav;
by SubRow SubCol RowwiSub ColwiSub;
run;
proc sort data=newc8tav;
by SubRow SubCol RowwiSub ColwiSub;
run;
proc sort data=newd8tav;
by SubRow SubCol RowwiSub ColwiSub;
run;
proc sort data=newa12tav;
by SubRow SubCol RowwiSub ColwiSub;
run;
proc sort data=newb12tav;
by SubRow SubCol RowwiSub ColwiSub;
run;
proc sort data=newc12tav;
by SubRow SubCol RowwiSub ColwiSub;
run;
proc sort data=newd12tav;
```

```
by SubRow SubCol RowwiSub ColwiSub;
run;

/* MERGED UNIGENE FILE WITH RAW DATA*/
data newa1tav2;
merge unigene newa1tav (drop = Row);
by SubRow SubCol RowwiSub ColwiSub;
run;
data newb1tav2;
merge unigene newb1tav (drop = Row);
by SubRow SubCol RowwiSub ColwiSub;
run;
data newc1tav2;
merge unigene newc1tav (drop = Row);
by SubRow SubCol RowwiSub ColwiSub;
run;
data newd1tav2;
merge unigene newd1tav (drop = Row);
by SubRow SubCol RowwiSub ColwiSub;
run;
data newa2tav2;
merge unigene newa2tav (drop = Row);
by SubRow SubCol RowwiSub ColwiSub;
run;
data newb2tav2;
merge unigene newb2tav (drop = Row);
by SubRow SubCol RowwiSub ColwiSub;
run;
data newc2tav2;
merge unigene newc2tav (drop = Row);
by SubRow SubCol RowwiSub ColwiSub;
run;
data newd2tav2;
merge unigene newd2tav (drop = Row);
by SubRow SubCol RowwiSub ColwiSub;
run;
data newa4tav2;
merge unigene newa4tav (drop = Row);
by SubRow SubCol RowwiSub ColwiSub;
run;
data newb4tav2;
merge unigene newb4tav (drop = Row);
by SubRow SubCol RowwiSub ColwiSub;
run;
data newc4tav2;
merge unigene newc4tav (drop = Row);
```

```
by SubRow SubCol RowwiSub ColwiSub;
run;
data newd4tav2;
merge unigene newd4tav (drop = Row);
by SubRow SubCol RowwiSub ColwiSub;
run;
data newa8tav2;
merge unigene newa8tav (drop = Row);
by SubRow SubCol RowwiSub ColwiSub;
run;
data newb8tav2;
merge unigene newb8tav (drop = Row);
by SubRow SubCol RowwiSub ColwiSub;
run;
data newc8tav2;
merge unigene newc8tav (drop = Row);
by SubRow SubCol RowwiSub ColwiSub;
run;
data newd8tav2;
merge unigene newd8tav (drop = Row);
by SubRow SubCol RowwiSub ColwiSub;
run;
data newa12tav2;
merge unigene newa12tav (drop = Row);
by SubRow SubCol RowwiSub ColwiSub;
run;
data newb12tav2;
merge unigene newb12tav (drop = Row);
by SubRow SubCol RowwiSub ColwiSub;
run;
data newc12tav2;
merge unigene newc12tav (drop = Row);
by SubRow SubCol RowwiSub ColwiSub;
run;
data newd12tav2;
merge unigene newd12tav (drop = Row);
by SubRow SubCol RowwiSub ColwiSub;
run;

/* REMOVED ROWS WITH MISSING SPOTS IN DATA*/
data newa1tav2;
set newa1tav2;
if SubRow>8 and RowwiSub>16 then delete;
if IntAa-TotalQCScore = '.' then delete;
run;
data newb1tav2;
```

```
set newb1tav2;
if SubRow>8 and RowwiSub>16 then delete;
if IntAb-TotalQCScore = '.' then delete;
run;
data newc1tav2;
set newc1tav2;
if SubRow>8 and RowwiSub>16 then delete;
if IntAc-TotalQCScore = '.' then delete;
run;
data newd1tav2;
set newd1tav2;
if SubRow>8 and RowwiSub>16 then delete;
if IntAd-TotalQCScore = '.' then delete;
run;
data newa2tav2;
set newa2tav2;
if SubRow>8 and RowwiSub>16 then delete;
if IntAa-TotalQCScore = '.' then delete;
run;
data newb2tav2;
set newb2tav2;
if SubRow>8 and RowwiSub>16 then delete;
if IntAb-TotalQCScore = '.' then delete;
run;
data newc2tav2;
set newc2tav2;
if SubRow>8 and RowwiSub>16 then delete;
if IntAc-TotalQCScore = '.' then delete;
run;
data newd2tav2;
set newd2tav2;
if SubRow>8 and RowwiSub>16 then delete;
if IntAd-TotalQCScore = '.' then delete;
run;
data newa4tav2;
set newa4tav2;
if SubRow>8 and RowwiSub>16 then delete;
if IntAa-TotalQCScore = '.' then delete;
run;
data newb4tav2;
set newb4tav2;
if SubRow>8 and RowwiSub>16 then delete;
if IntAb-TotalQCScore = '.' then delete;
run;
data newc4tav2;
set newc4tav2;
```

```
if SubRow>8 and RowwiSub>16 then delete;
if IntAc-TotalQCScore = '.' then delete;
run;
data newd4tav2;
set newd4tav2;
if SubRow>8 and RowwiSub>16 then delete;
if IntAd-TotalQCScore = '.' then delete;
run;
data newa8tav2;
set newa8tav2;
if SubRow>8 and RowwiSub>16 then delete;
if IntAa-TotalQCScore = '.' then delete;
run;
data newb8tav2;
set newb8tav2;
if SubRow>8 and RowwiSub>16 then delete;
if IntAb-TotalQCScore = '.' then delete;
run;
data newc8tav2;
set newc8tav2;
if SubRow>8 and RowwiSub>16 then delete;
if IntAc-TotalQCScore = '.' then delete;
run;
data newd8tav2;
set newd8tav2;
if SubRow>8 and RowwiSub>16 then delete;
if IntAd-TotalQCScore = '.' then delete;
run;
data newa12tav2;
set newa12tav2;
if SubRow>8 and RowwiSub>16 then delete;
if IntAa-TotalQCScore = '.' then delete;
run;
data newb12tav2;
set newb12tav2;
if SubRow>8 and RowwiSub>16 then delete;
if IntAb-TotalQCScore = '.' then delete;
run;
data newc12tav2;
set newc12tav2;
if SubRow>8 and RowwiSub>16 then delete;
if IntAc-TotalQCScore = '.' then delete;
run;
data newd12tav2;
set newd12tav2;
if SubRow>8 and RowwiSub>16 then delete;
```

```
if IntAd-TotalQCScore = '.' then delete;
run;

/*CREATED PRINT TIP FIELD*/
data newa1tav2;
set newa1tav2;
if SubRow<9 and RowwiSub<18 then print=1;
if SubRow<9 and RowwiSub>17 then print=2;
if SubRow>8 then print=1;
run;
data newb1tav2;
set newb1tav2;
if SubRow<9 and RowwiSub<18 then print=1;
if SubRow<9 and RowwiSub>17 then print=2;
if SubRow>8 then print=1;
run;
data newc1tav2;
set newc1tav2;
if SubRow<9 and RowwiSub<18 then print=1;
if SubRow<9 and RowwiSub>17 then print=2;
if SubRow>8 then print=1;
run;
data newd1tav2;
set newd1tav2;
if SubRow<9 and RowwiSub<18 then print=1;
if SubRow<9 and RowwiSub>17 then print=2;
if SubRow>8 then print=1;
run;
data newa2tav2;
set newa2tav2;
if SubRow<9 and RowwiSub<18 then print=1;
if SubRow<9 and RowwiSub>17 then print=2;
if SubRow>8 then print=1;
run;
data newb2tav2;
set newb2tav2;
if SubRow<9 and RowwiSub<18 then print=1;
if SubRow<9 and RowwiSub>17 then print=2;
if SubRow>8 then print=1;
run;
data newc2tav2;
set newc2tav2;
if SubRow<9 and RowwiSub<18 then print=1;
if SubRow<9 and RowwiSub>17 then print=2;
if SubRow>8 then print=1;
run;
```

```
data newd2tav2;
set newd2tav2;
if SubRow<9 and RowwiSub<18 then print=1;
if SubRow<9 and RowwiSub>17 then print=2;
if SubRow>8 then print=1;
run;
data newa4tav2;
set newa4tav2;
if SubRow<9 and RowwiSub<18 then print=1;
if SubRow<9 and RowwiSub>17 then print=2;
if SubRow>8 then print=1;
run;
data newb4tav2;
set newb4tav2;
if SubRow<9 and RowwiSub<18 then print=1;
if SubRow<9 and RowwiSub>17 then print=2;
if SubRow>8 then print=1;
run;
data newc4tav2;
set newc4tav2;
if SubRow<9 and RowwiSub<18 then print=1;
if SubRow<9 and RowwiSub>17 then print=2;
if SubRow>8 then print=1;
run;
data newd4tav2;
set newd4tav2;
if SubRow<9 and RowwiSub<18 then print=1;
if SubRow<9 and RowwiSub>17 then print=2;
if SubRow>8 then print=1;
run;
data newa8tav2;
set newa8tav2;
if SubRow<9 and RowwiSub<18 then print=1;
if SubRow<9 and RowwiSub>17 then print=2;
if SubRow>8 then print=1;
run;
data newb8tav2;
set newb8tav2;
if SubRow<9 and RowwiSub<18 then print=1;
if SubRow<9 and RowwiSub>17 then print=2;
if SubRow>8 then print=1;
run;
data newc8tav2;
set newc8tav2;
if SubRow<9 and RowwiSub<18 then print=1;
if SubRow<9 and RowwiSub>17 then print=2;
```

```
if SubRow>8 then print=1;
run;
data newd8tav2;
set newd8tav2;
if SubRow<9 and RowwiSub<18 then print=1;
if SubRow<9 and RowwiSub>17 then print=2;
if SubRow>8 then print=1;
run;
data newa12tav2;
set newa12tav2;
if SubRow<9 and RowwiSub<18 then print=1;
if SubRow<9 and RowwiSub>17 then print=2;
if SubRow>8 then print=1;
run;
data newb12tav2;
set newb12tav2;
if SubRow<9 and RowwiSub<18 then print=1;
if SubRow<9 and RowwiSub>17 then print=2;
if SubRow>8 then print=1;
run;
data newc12tav2;
set newc12tav2;
if SubRow<9 and RowwiSub<18 then print=1;
if SubRow<9 and RowwiSub>17 then print=2;
if SubRow>8 then print=1;
run;
data newd12tav2;
set newd12tav2;
if SubRow<9 and RowwiSub<18 then print=1;
if SubRow<9 and RowwiSub>17 then print=2;
if SubRow>8 then print=1;
run;

/* VSN FILE CREATION FOR STANDARDIZATION PROCEDURE*/
data vsntime1;
merge newa1tav2 (keep = SubRow SubCol RowwiSub ColwiSub IntAa IntBa print)
newb1tav2 (keep=SubRow SubCol RowwiSub ColwiSub IntAb IntBb print) newc1tav2
(keep = SubRow SubCol RowwiSub ColwiSub IntAc IntBc print) newd1tav2 (keep =
SubRow SubCol RowwiSub ColwiSub IntAd IntBd print) ;
by SubRow SubCol RowwiSub ColwiSub;
run;
data vsntime2;
merge newa2tav2 (keep = SubRow SubCol RowwiSub ColwiSub IntAa IntBa print)
newb2tav2 (keep=SubRow SubCol RowwiSub ColwiSub IntAb IntBb print) newc2tav2
(keep = SubRow SubCol RowwiSub ColwiSub IntAc IntBc print) newd2tav2 (keep =
SubRow SubCol RowwiSub ColwiSub IntAd IntBd print) ;
```

```
by SubRow SubCol RowwiSub ColwiSub;
run;
data vsntime3;
merge newa4tav2 (keep = SubRow SubCol RowwiSub ColwiSub IntAa IntBa print)
newb4tav2 (keep=SubRow SubCol RowwiSub ColwiSub IntAb IntBb print) newc4tav2
(keep = SubRow SubCol RowwiSub ColwiSub IntAc IntBc print) newd4tav2 (keep =
SubRow SubCol RowwiSub ColwiSub IntAd IntBd print) ;
by SubRow SubCol RowwiSub ColwiSub;
run;
data vsntime4;
merge newa8tav2 (keep = SubRow SubCol RowwiSub ColwiSub IntAa IntBa print)
newb8tav2 (keep=SubRow SubCol RowwiSub ColwiSub IntAb IntBb print) newc8tav2
(keep = SubRow SubCol RowwiSub ColwiSub IntAc IntBc print) newd8tav2 (keep =
SubRow SubCol RowwiSub ColwiSub IntAd IntBd print) ;
by SubRow SubCol RowwiSub ColwiSub;
run;
data vsntime5;
merge newa12tav2 (keep = SubRow SubCol RowwiSub ColwiSub IntAa IntBa print)
newb12tav2 (keep=SubRow SubCol RowwiSub ColwiSub IntAb IntBb print)
newc12tav2 (keep = SubRow SubCol RowwiSub ColwiSub IntAc IntBc print)
newd12tav2 (keep = SubRow SubCol RowwiSub ColwiSub IntAd IntBd print) ;
by SubRow SubCol RowwiSub ColwiSub;
run;

/* DELETED INTENSITIES FOR SLIDES MISSING TWO OR MORE VALUES*/
data expvsntime1;
set vsntime1;
if IntAa = "." and IntBa="." and IntAb="." and IntBb="." and IntAc="." and IntBc="."
and IntAd="." and IntBd="." then delete;
run;
data expvsntime2;
set vsntime2;
if IntAa = "." and IntBa="." and IntAb="." and IntBb="." and IntAc="." and IntBc="."
and IntAd="." and IntBd="." then delete;
run;
data expvsntime3;
set vsntime3;
if IntAa = "." and IntBa="." and IntAb="." and IntBb="." and IntAc="." and IntBc="."
and IntAd="." and IntBd="." then delete;
run;
data expvsntime4;
set vsntime4;
if IntAa = "." and IntBa="." and IntAb="." and IntBb="." and IntAc="." and IntBc="."
and IntAd="." and IntBd="." then delete;
run;
data expvsntime5;
```

```
set vsntime5;
if IntAa = "." and IntBa="." and IntAb="." and IntBb="." and IntAc="." and IntBc="."
and IntAd="." and IntBd="." then delete;
run;

/*INCORPORATED TIME, ARRAY, TRT, DYE, INTENSITY*/
data newform1;
set expvsntime1;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=1;array=1;trt='c';dye='g';intensity=IntAa;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=1;array=1;trt='t';dye='r';intensity=IntBa;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=1;array=2;trt='c';dye='g';intensity=IntAb;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=1;array=2;trt='t';dye='r';intensity=IntBb;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=1;array=3;trt='t';dye='g';intensity=IntAc;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=1;array=3;trt='c';dye='r';intensity=IntBc;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=1;array=4;trt='t';dye='g';intensity=IntAd;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=1;array=4;trt='c';dye='r';intensity=IntBd;output;
keep SubRow SubCol RowwiSub ColwiSub print time array trt dye intensity;
run;
data newform2;
set expvsntime2;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=2;array=1;trt='c';dye='g';intensity=IntAa;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=2;array=1;trt='t';dye='r';intensity=IntBa;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=2;array=2;trt='c';dye='g';intensity=IntAb;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=2;array=2;trt='t';dye='r';intensity=IntBb;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=2;array=3;trt='t';dye='g';intensity=IntAc;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=2;array=3;trt='c';dye='r';intensity=IntBc;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=2;array=4;trt='t';dye='g';intensity=IntAd;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=2;array=4;trt='c';dye='r';intensity=IntBd;output;
keep SubRow SubCol RowwiSub ColwiSub print time array trt dye intensity;
run;
```

```
data newform3;
set expvsntime3;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=3;array=1;trt='c';dye='g';intensity=IntAa;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=3;array=1;trt='t';dye='r';intensity=IntBa;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=3;array=2;trt='c';dye='g';intensity=IntAb;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=3;array=2;trt='t';dye='r';intensity=IntBb;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=3;array=3;trt='t';dye='g';intensity=IntAc;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=3;array=3;trt='c';dye='r';intensity=IntBc;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=3;array=4;trt='t';dye='g';intensity=IntAd;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=3;array=4;trt='c';dye='r';intensity=IntBd;output;
keep SubRow SubCol RowwiSub ColwiSub print time array trt dye intensity;
run;
data newform4;
set expvsntime4;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=4;array=1;trt='c';dye='g';intensity=IntAa;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=4;array=1;trt='t';dye='r';intensity=IntBa;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=4;array=2;trt='c';dye='g';intensity=IntAb;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=4;array=2;trt='t';dye='r';intensity=IntBb;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=4;array=3;trt='t';dye='g';intensity=IntAc;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=4;array=3;trt='c';dye='r';intensity=IntBc;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=4;array=4;trt='t';dye='g';intensity=IntAd;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=4;array=4;trt='c';dye='r';intensity=IntBd;output;
keep SubRow SubCol RowwiSub ColwiSub print time array trt dye intensity;
run;
data newform5;
set expvsntime5;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=5;array=1;trt='c';dye='g';intensity=IntAa;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=5;array=1;trt='t';dye='r';intensity=IntBa;output;
```

```
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=5;array=2;trt='c';dye='g';intensity=IntAb;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=5;array=2;trt='t';dye='r';intensity=IntBb;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=5;array=3;trt='t';dye='g';intensity=IntAc;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=5;array=3;trt='c';dye='r';intensity=IntBc;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=5;array=4;trt='t';dye='g';intensity=IntAd;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;print=
print;time=5;array=4;trt='c';dye='r';intensity=IntBd;output;
keep SubRow SubCol RowwiSub ColwiSub print time array trt dye intensity;
run;

/*CREATION OF SEPARATE DATA SETS BY PRINT TIP FOR PRINT TIP
DEPENDENT STANDARDIZATION*/
data t1_p1;
set newform1;
if print=2 then delete;
run;

data t1_p2;
set newform1;
if print=1 then delete;
run;

data t2_p1;
set newform2;
if print=2 then delete;
run;

data t2_p2;
set newform2;
if print=1 then delete;
run;
data t3_p1;
set newform3;
if print=2 then delete;
run;
data t3_p2;
set newform3;
if print=1 then delete;
run;
data t4_p1;
set newform4;
```

```
if print=2 then delete;
run;

data t4_p2;
set newform4;
if print=1 then delete;
run;
data t5_p1;
set newform5;
if print=2 then delete;
run;
data t5_p2;
set newform5;
if print=1 then delete;
run;

/*MERGED ALL NEWFORM DATA SETS*/
data finalnewform;
set newform1 newform2 newform3 newform4 newform5;
run;

proc sort data=finalnewform;
by SubRow SubCol RowwiSub ColwiSub;
run;

/*MERGE FINALNEWFORM WITH GENE NAMES*/
data timecourse;
merge unigene finalnewform;
by SubRow SubCol RowwiSub ColwiSub;
keep Gene TUC_ID time array trt dye intensity SubRow SubCol RowwiSub ColwiSub
print;
run;

proc sort data = timecourse;
by gene;
run;

/*CREATED STATEMENT BECAUSE TIMECOURSE WAS TOO BIG TO EXPORT
TO EXCEL*/
libname trial'C:\Carletha';
data trial.timecourse;
set timecourse;
run;

/*CREATION OF FILES FOR VSN IN BIOCONDUCTOR IN R*/
data vsntime1new;
```

```
infile 'C:\Carletha\Microarray Information\vsntime1_1.csv' dlm = ",";
input SubRow SubCol RowwiSub ColwiSub IntAa1 IntBa1 IntAb1 IntBb1 IntAc1
IntBc1 IntAd1 IntBd1 ;
run;
proc sort data = vsntime1new;
by SubRow SubCol RowwiSub ColwiSub;
run;
data vsntime2new;
infile 'C:\Carletha\Microarray Information\vsntime2_1.csv' dlm = ",";
input SubRow SubCol RowwiSub ColwiSub IntAa2 IntBa2 IntAb2 IntBb2 IntAc2
IntBc2 IntAd2 IntBd2 ;
run;
proc sort data = vsntime2new;
by SubRow SubCol RowwiSub ColwiSub;
run;
data vsntime3new;
infile 'C:\Carletha\Microarray Information\vsntime3_1.csv' dlm = ",";
input SubRow SubCol RowwiSub ColwiSub IntAa3 IntBa3 IntAb3 IntBb3 IntAc3
IntBc3 IntAd3 IntBd3;
run;
proc sort data = vsntime3new;
by SubRow SubCol RowwiSub ColwiSub;
run;
data vsntime4new;
infile 'C:\Carletha\Microarray Information\vsntime4_1.csv' dlm = ",";
input SubRow SubCol RowwiSub ColwiSub IntAa4 IntBa4 IntAb4 IntBb4 IntAc4
IntBc4 IntAd4 IntBd4;
run;
proc sort data = vsntime4new;
by SubRow SubCol RowwiSub ColwiSub;
run;
data vsntime5new;
infile 'C:\Carletha\Microarray Information\vsntime5_1.csv' dlm = ",";
input SubRow SubCol RowwiSub ColwiSub IntAa5 IntBa5 IntAb5 IntBb5 IntAc5
IntBc5 IntAd5 IntBd5;
run;
proc sort data = vsntime5new;
by SubRow SubCol RowwiSub ColwiSub;
run;

data multiplevsn;
merge vsntime1new vsntime2new vsntime3new vsntime4new vsntime5new;
by SubRow SubCol RowwiSub ColwiSub;
run;
```

```
/*DELETED PROFILES WITH MISSING INTENSITIES BECAUSE
STANDARDIZATION PROCEDURER WILL NOT ACCEPT*/
data multiplevsnforR;
set multiplevsn;
if IntAa1-IntBd1 = "." then delete;
if IntAa2-IntBd2 = "." then delete;
if IntAa3-IntBd3 = "." then delete;
if IntAa4-IntBd4 = "." then delete;
if IntAa5-IntBd5 = "." then delete;
run;

/*CREATED PRINTTIP FOR R STRATA COMMAND FOR STANDARDIZATION*/
data printtip;
set vsntime1 (keep = SubRow SubCol RowwiSub ColwiSub print);
run;

/*MERGED PRINTTIP AND INTENSITIES FOR R STANDARDIZATION*/
data printandvsn;
merge multiplevsnforR printtip;
by SubRow SubCol RowwiSub ColwiSub;
run;

data printandvsnfinal;
set printandvsn;
if IntAa1-IntBd1 = "." then delete;
if IntAa2-IntBd2 = "." then delete;
if IntAa3-IntBd3 = "." then delete;
if IntAa4-IntBd4 = "." then delete;
if IntAa5-IntBd5 = "." then delete;
run;

/*STANDARDIZED DATA READ INTO SAS*/
data unedit_povsntc;
infile 'C:\Carletha\postvsntimecourse.csv' dlm = "," dsd truncover lrecl=10000;
input SubRow SubCol RowwiSub ColwiSub IntAa1 IntBa1 IntAb1 IntBb1 IntAc1 IntBc1
IntAd1 IntBd1 IntAa2 IntBa2 IntAb2 IntBb2 IntAc2 IntBc2 IntAd2 IntBd2 IntAa3 IntBa3
IntAb3 IntBb3 IntAc3 IntBc3 IntAd3 IntBd3 IntAa4 IntBa4 IntAb4 IntBb4 IntAc4 IntBc4
        IntAd4 IntBd4 IntAa5 IntBa5 IntAb5 IntBb5 IntAc5 IntBc5 IntAd5 IntBd5;
run;

/* REMOVED ARRAY #6 FROM ANALYSIS DUE TO BAD M VS A PLOT*/
data povsntc;
set unedit_povsntc;
drop IntAb2 IntBb2;
run;
```

```
data unigene;
infile 'C:\Carletha\unigene.csv' dlm = "," dsd truncover lrecl=10000 firstobs=2;
input SubRow SubCol RowwiSub ColwiSub Gene   Gi :  10. TUC_ID :  30. Homo1_PID
:  80. Homo1_Desc :  80. Homo1_Species :  80. Homo2_PID :  80.
Homo2_Desc :  80. Homo2_Species :  80. Homo3_PID :  80. Homo3_Desc :  80.
Homo3_Species :  80.;
if gene='BLANK' then delete;
if gene='blank' then delete;
if gene='xxxx' then delete;
if gene='Blank' then delete;
run;

proc sort data = unigene;
by SubRow SubCol RowwiSub ColwiSub;
run;

/*DATA IN NECESSARY FORMAT WITH NEEDED INFO TO RUN GENERAL
LINEAR MODEL*/
data tcnewform;
set povsntc;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=1
;array=1;trt='c';dye='g';intensity=IntAa1;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=1
;array=1;trt='t';dye='r';intensity=IntBa1;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=1
;array=2;trt='c';dye='g';intensity=IntAb1;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=1
;array=2;trt='t';dye='r';intensity=IntBb1;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=1
;array=3;trt='t';dye='g';intensity=IntAc1;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=1
;array=3;trt='c';dye='r';intensity=IntBc1;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=1
;array=4;trt='t';dye='g';intensity=IntAd1;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=1
;array=4;trt='c';dye='r';intensity=IntBd1;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=2
;array=1;trt='c';dye='g';intensity=IntAa2;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=2
;array=1;trt='t';dye='r';intensity=IntBa2;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=2
;array=3;trt='t';dye='g';intensity=IntAc2;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=2
;array=3;trt='c';dye='r';intensity=IntBc2;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=2
;array=4;trt='t';dye='g';intensity=IntAd2;output;
```

SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=2
;array=4;trt='c';dye='r';intensity=IntBd2;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=3
;array=1;trt='c';dye='g';intensity=IntAa3;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=3
;array=1;trt='t';dye='r';intensity=IntBa3;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=3
;array=2;trt='c';dye='g';intensity=IntAb3;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=3
;array=2;trt='t';dye='r';intensity=IntBb3;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=3
;array=3;trt='t';dye='g';intensity=IntAc3;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=3
;array=3;trt='c';dye='r';intensity=IntBc3;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=3
;array=4;trt='t';dye='g';intensity=IntAd3;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=3
;array=4;trt='c';dye='r';intensity=IntBd3;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=4
;array=1;trt='c';dye='g';intensity=IntAa4;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=4
;array=1;trt='t';dye='r';intensity=IntBa4;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=4
;array=2;trt='c';dye='g';intensity=IntAb4;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=4
;array=2;trt='t';dye='r';intensity=IntBb4;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=4
;array=3;trt='t';dye='g';intensity=IntAc4;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=4
;array=3;trt='c';dye='r';intensity=IntBc4;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=4
;array=4;trt='t';dye='g';intensity=IntAd4;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=4
;array=4;trt='c';dye='r';intensity=IntBd4;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=5
;array=1;trt='c';dye='g';intensity=IntAa5;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=5
;array=1;trt='t';dye='r';intensity=IntBa5;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=5
;array=2;trt='c';dye='g';intensity=IntAb5;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=5
;array=2;trt='t';dye='r';intensity=IntBb5;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=5
;array=3;trt='t';dye='g';intensity=IntAc5;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=5
;array=3;trt='c';dye='r';intensity=IntBc5;output;

```
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=5
;array=4;trt='t';dye='g';intensity=IntAd5;output;
SubRow=SubRow;SubCol=SubCol;RowwiSub=RowwiSub;ColwiSub=ColwiSub;time=5
;array=4;trt='c';dye='r';intensity=IntBd5;output;
keep SubRow SubCol RowwiSub ColwiSub time array trt dye intensity;
run;

data timecourse1a;
merge unigene tcnewform;
by SubRow SubCol RowwiSub ColwiSub;
keep Gene TUC_ID time array trt dye intensity SubRow SubCol RowwiSub ColwiSub;
run;

/*REMOVED ROWS FOR WHICH NO SPOTS WERE PRESENT ON ORIGINAL
MICROARRAY*/
data timecourse1b;
set timecourse1a;
if intensity = "." then delete;
if SubRow>8 and RowwiSub>16 then delete;
run;

proc sort data =timecourse1b;
by descending Gene;
run;
/*CREATED DATA SETS BY TIME TO RUN MODELS SEPERATELY*/
data timecourse1;
set timecourse1b;
keep Gene TUC_ID time array trt dye intensity SubRow SubCol RowwiSub ColwiSub;
if time=2 then delete;
if time=3 then delete;
if time=4 then delete;
if time=5 then delete;
run;
data timecourse2;
set timecourse1b;
keep Gene TUC_ID time array trt dye intensity SubRow SubCol RowwiSub ColwiSub;
if time=1 then delete;
if time=3 then delete;
if time=4 then delete;
if time=5 then delete;
run;
data timecourse3;
set timecourse1b;
keep Gene TUC_ID time array trt dye intensity SubRow SubCol RowwiSub ColwiSub;
if time=1 then delete;
if time=2 then delete;
```

```
if time=4 then delete;
if time=5 then delete;
run;
data timecourse4;
set timecourse1b;
keep Gene TUC_ID time array trt dye intensity SubRow SubCol RowwiSub ColwiSub;
if time=1 then delete;
if time=2 then delete;
if time=3 then delete;
if time=5 then delete;
run;
data timecourse5;
set timecourse1b;
keep Gene TUC_ID time array trt dye intensity SubRow SubCol RowwiSub ColwiSub;
if time=1 then delete;
if time=2 then delete;
if time=3 then delete;
if time=4 then delete;
run;

/* CREATED TIMECOURSE_REV TO DROP TUC_ID BECAUSE MODEL DIDN'T
LIKE IT*/
data timecourse1_rev;
set timecourse1;
drop TUC_ID;
run;
proc sort data = timecourse1_rev;
by descending Gene;
run;
data timecourse2_rev;
set timecourse2;
drop TUC_ID;
run;
proc sort data = timecourse2_rev;
by descending Gene;
run;
data timecourse3_rev;
set timecourse3;
drop TUC_ID;
run;
proc sort data = timecourse3_rev;
by descending Gene;
run;
data timecourse4_rev;
set timecourse4;
drop TUC_ID;
```

```
run;
proc sort data = timecourse4_rev;
by descending Gene;
run;
data timecourse5_rev;
set timecourse5;
drop TUC_ID;
run;
proc sort data = timecourse5_rev;
by descending Gene;
run;

/*THIS IS WHERE THE MODEL STARTS*/
/*THIS STEP IS TO KEEP THE LOG FROM FILLING UP*/
options nonotes;
ods listing close;

/*GENMOD WITH NORMAL DISTRIBUTION*/
proc genmod data = timecourse1_rev;
class dye array trt Gene;
     model intensity = dye array trt/dist=normal type3;
by descending Gene;
make 'type3' out=p_table1n;
run;
data p_value1n;
set p_table1n;
if source='trt' then p_value=ProbChiSq;
if mod(_N_,4)=0 then output;
keep Gene p_value;
run;

proc genmod data = timecourse2_rev;
class dye array trt Gene;
     model intensity = dye array trt/dist=normal type3;
by descending Gene;
make 'type3' out=p_table2n;
run;
data p_value2n;
set p_table2n;
if source='trt' then p_value=ProbChiSq;
if mod(_N_,4)=0 then output;
keep Gene p_value;
run;

proc genmod data = timecourse3_rev;
class dye array trt Gene;
```

```
     model intensity = dye array trt/dist=normal type3;
by descending Gene;
make 'type3' out=p_table3n;
run;
data p_value3n;
set p_table3n;
if source='trt' then p_value=ProbChiSq;
if mod(_N_,4)=0 then output;
keep Gene p_value;
run;

proc genmod data = timecourse4_rev;
class dye array trt Gene;
     model intensity = dye array trt/dist=normal type3;
by descending Gene;
make 'type3' out=p_table4n;
run;
data p_value4n;
set p_table4n;
if source='trt' then p_value=ProbChiSq;
if mod(_N_,4)=0 then output;
keep Gene p_value;
run;

proc genmod data = timecourse5_rev;
class dye array trt Gene;
     model intensity = dye array trt/dist=normal type3;
by descending Gene;
make 'type3' out=p_table5n;
run;
data p_value5n;
set p_table5n;
if source='trt' then p_value=ProbChiSq;
if mod(_N_,4)=0 then output;
keep Gene p_value;
run;

/*RAN MODEL WITH ALL TIMES TOGETHER*/
proc genmod data = timecourse1b;
class dye array trt Gene time;
     model intensity = dye array trt time trt*time/dist=normal type3;
by descending Gene;
make 'type3' out=p_tableall5;
run;
data p_valueall5;
set p_tableall5;
```

```
if source='trt' then p_value=ProbChiSq;
if mod(_N_,4)=0 then output;
keep Gene p_value;
run;

options notes;
run;

/*CREATED DATA SETS FOR SOURCE=TRT ONLY, DELETED PROB. ='<.0001'*/
data t1_modelresults;
set p_table1n;
if Source='array' then delete;
if Source='dye' then delete;
if ProbChiSq=<.0001 then delete;
run;
data t2_modelresults;
set p_table2n;
if Source='array' then delete;
if Source='dye' then delete;
if ProbChiSq=<.0001 then delete;
run;
data t3_modelresults;
set p_table3n;
if Source='array' then delete;
if Source='dye' then delete;
if ProbChiSq=<.0001 then delete;
run;
data t4_modelresults;
set p_table4n;
if Source='array' then delete;
if Source='dye' then delete;
if ProbChiSq=<.0001 then delete;
run;
data t5_modelresults;
set p_table5n;
if Source='array' then delete;
if Source='dye' then delete;
if ProbChiSq=<.0001 then delete;
run;
data all5_modelresults;
set P_tableall5;
if Source ='array' then delete;
if Source ='dye' then delete;
if Source ='trt' then delete;
if Source ='time' then delete;
if ProbChiSq=<.0001 then delete;
```

```
run;

/*READ IN QOUTPUT TO COMPARE SIGNIFICANT LISTS*/
data genelisttime1;
infile 'C:\Carletha\Q Results (final)\fdrlists_t1.csv' dsd truncover lrecl=10000 firstobs=4;
input pvalue1 qvalue gene $ one1 five1;
run;
data genelisttime2;
infile 'C:\Carletha\Q Results (final)\fdrlists_t2.csv' dsd truncover lrecl=10000 firstobs=4;
input pvalue2 qvalue gene $ one2 five2;
run;
data genelisttime3;
infile 'C:\Carletha\Q Results (final)\fdrlists_t3.csv' dsd truncover lrecl=10000 firstobs=4;
input pvalue3 qvalue gene $ one3 five3;
run;
data genelisttime4;
infile 'C:\Carletha\Q Results (final)\fdrlists_t4.csv' dsd truncover lrecl=10000 firstobs=4;
input pvalue4 qvalue gene $ one4 five4;
run;
data genelisttime5;
infile 'C:\Carletha\Q Results (final)\fdrlists_t5.csv' dsd truncover lrecl=10000 firstobs=4;
input pvalue5 qvalue gene $ one5 five5;
run;

data list1;
set genelisttime1;
if one1='1' then output;
run;

data list2;
set genelisttime2;
if one2='1' then output;
run;

data list3;
set genelisttime3;
if one3='1' then output;
run;

data list4;
set genelisttime4;
if one4='1' then output;
run;

data list5;
set genelisttime5;
```

```
if one5='1' then output;
run;

/*GENES SIGNIFICANTLY DIFFERENT ACROSS ALL FIVE TIMES FOR A 1% FDR*/
data genelistall5;
infile 'C:\Carletha\Q Results (final)\fdrlists_all5.csv' dsd truncover lrecl=10000 firstobs=2;
input pvalue qvalue gene $ oneall5 fiveall5;
run;

/*LISTS FOR WHICH GENES ARE SIGNIFICANT FOR A GIVEN TIME FOR A 1% FDR*/
data onefdrt1;
set genelisttime1;
keep gene pvalue1 one1;
if one1='1' then output;
run;
data onefdrt2;
set genelisttime2;
keep gene pvalue2 one2;
if one2='1' then output;
run;
data onefdrt3;
set genelisttime3;
keep gene pvalue3 one3;
if one3='1' then output;
run;
data onefdrt4;
set genelisttime4;
keep gene pvalue4 one4;
if one4='1' then output;
run;
data onefdrt5;
set genelisttime5;
keep gene pvalue5 one5;
if one5='1' then output;
run;

/*SORTED DATA FOR MERGE*/
proc sort data=onefdrt1;
by gene;
run;
proc sort data=onefdrt2;
by gene;
run;
```

```
proc sort data=onefdrt3;
by gene;
run;
proc sort data=onefdrt4;
by gene;
run;
proc sort data=onefdrt5;
by gene;
run;
proc sort data=genelistall5;
by gene;
run;

/*ELEMENTARY LIST COMPARISONS FOR A 1% FDR*/
data fdr1t1andt2;
merge onefdrt1 onefdrt2;
by gene;
if pvalue1='.' then delete;
if pvalue2='.' then delete;
run;
data fdr1t1andt3;
merge onefdrt1 onefdrt3;
by gene;
if pvalue1='.' then delete;
if pvalue3='.' then delete;
run;
data fdr1t1andt4;
merge onefdrt1 onefdrt4;
by gene;
if pvalue1='.' then delete;
if pvalue4='.' then delete;
run;
data fdr1t1andt5;
merge onefdrt1 onefdrt5;
by gene;
if pvalue1='.' then delete;
if pvalue5='.' then delete;
run;
data fdr1t2andt3;
merge onefdrt2 onefdrt3;
by gene;
if pvalue2='.' then delete;
if pvalue3='.' then delete;
run;
data fdr1t2andt4;
merge onefdrt2 onefdrt4;
```

```
by gene;
if pvalue2='.' then delete;
if pvalue4='.' then delete;
run;
data fdr1t2andt5;
merge onefdrt2 onefdrt5;
by gene;
if pvalue2='.' then delete;
if pvalue5='.' then delete;
run;
data fdr1t3andt4;
merge onefdrt3 onefdrt4;
by gene;
if pvalue3='.' then delete;
if pvalue4='.' then delete;
run;
data fdr1t4andt5;
merge onefdrt4 onefdrt5;
by gene;
if pvalue4='.' then delete;
if pvalue5='.' then delete;
run;

/*THESE ARE THE LISTS FOUND TO HAVE A SIGNIFICANT DIFFERENCE B/W
TREATMENT AND CONTROL INTENSITIES W/ 1% FDR*/
data merge_cluster;
merge onefdrt1 onefdrt2 onefdrt3 onefdrt4 onefdrt5;
by gene;
run;

/*CREATED DATA SETS TO IDENTIFY ALL POSSIBLE 2-WAY,3,4,5 WAY
COMPARISONS*/
data pairwise1;
set merge_cluster;
if one1='.' then delete;
if one2='.' then delete;
if one3='1' then delete;
if one4='1' then delete;
if one5='1' then delete;
run;

data pairwise2;
set merge_cluster;
if one1='.' then delete;
if one3='.' then delete;
if one2='1' then delete;
```

```
if one4='1' then delete;
if one5='1' then delete;
run;

data pairwise3;
set merge_cluster;
if one1='.' then delete;
if one4='.' then delete;
if one2='1' then delete;
if one3='1' then delete;
if one5='1' then delete;
run;

data pairwise4;
set merge_cluster;
if one1='.' then delete;
if one5='.' then delete;
if one2='1' then delete;
if one3='1' then delete;
if one4='1' then delete;
run;

data pairwise5;
set merge_cluster;
if one2='.' then delete;
if one3='.' then delete;
if one1='1' then delete;
if one4='1' then delete;
if one5='1' then delete;
run;

data pairwise6;
set merge_cluster;
if one2='.' then delete;
if one4='.' then delete;
if one1='1' then delete;
if one3='1' then delete;
if one5='1' then delete;
run;

data pairwise7;
set merge_cluster;
if one2='.' then delete;
if one5='.' then delete;
if one1='1' then delete;
if one3='1' then delete;
```

```
if one4='1' then delete;
run;

data pairwise8;
set merge_cluster;
if one3='.' then delete;
if one4='.' then delete;
if one1='1' then delete;
if one2='1' then delete;
if one5='1' then delete;
run;

data pairwise9;
set merge_cluster;
if one3='.' then delete;
if one5='.' then delete;
if one1='1' then delete;
if one2='1' then delete;
if one4='1' then delete;
run;

data pairwise10;
set merge_cluster;
if one4='.' then delete;
if one5='.' then delete;
if one1='1' then delete;
if one2='1' then delete;
if one3='1' then delete;
run;

data triplet1;
set merge_cluster;
if one1='.' then delete;
if one2='.' then delete;
if one3='.' then delete;
if one4='1' then delete;
if one5='1' then delete;
run;

data triplet2;
set merge_cluster;
if one1='.' then delete;
if one2='.' then delete;
if one4='.' then delete;
if one3='1' then delete;
if one5='1' then delete;
```

```
run;

data triplet3;
set merge_cluster;
if one1='.' then delete;
if one2='.' then delete;
if one5='.' then delete;
if one3='1' then delete;
if one4='1' then delete;
run;

data triplet4;
set merge_cluster;
if one1='.' then delete;
if one3='.' then delete;
if one4='.' then delete;
if one2='1' then delete;
if one5='1' then delete;
run;

data triplet5;
set merge_cluster;
if one1='.' then delete;
if one3='.' then delete;
if one5='.' then delete;
if one2='1' then delete;
if one4='1' then delete;
run;

data triplet6;
set merge_cluster;
if one1='.' then delete;
if one4='.' then delete;
if one5='.' then delete;
if one2='1' then delete;
if one3='1' then delete;
run;

data triplet7;
set merge_cluster;
if one2='.' then delete;
if one3='.' then delete;
if one4='.' then delete;
if one1='1' then delete;
if one5='1' then delete;
run;
```

```
data triplet8;
set merge_cluster;
if one2='.' then delete;
if one3='.' then delete;
if one5='.' then delete;
if one1='1' then delete;
if one4='1' then delete;
run;

data triplet9;
set merge_cluster;
if one3='.' then delete;
if one4='.' then delete;
if one5='.' then delete;
if one1='1' then delete;
if one2='1' then delete;
run;

data triplet10;
set merge_cluster;
if one2='.' then delete;
if one4='.' then delete;
if one5='.' then delete;
if one1='1' then delete;
if one3='1' then delete;
run;

data quad1;
set merge_cluster;
if one1='.' then delete;
if one2='.' then delete;
if one3='.' then delete;
if one4='.' then delete;
if one5='1' then delete;
run;

data quad2;
set merge_cluster;
if one1='.' then delete;
if one2='.' then delete;
if one3='.' then delete;
if one5='.' then delete;
if one4='1' then delete;
run;
```

```
data quad3;
set merge_cluster;
if one1='.' then delete;
if one2='.' then delete;
if one4='.' then delete;
if one5='.' then delete;
if one3='1' then delete;
run;

data quad4;
set merge_cluster;
if one1='.' then delete;
if one3='.' then delete;
if one4='.' then delete;
if one5='.' then delete;
if one2='1' then delete;
run;

data quad5;
set merge_cluster;
if one2='.' then delete;
if one3='.' then delete;
if one4='.' then delete;
if one5='.' then delete;
if one1='1' then delete;
run;

data quint;
set merge_cluster;
if one1='.' then delete;
if one2='.' then delete;
if one3='.' then delete;
if one4='.' then delete;
if one5='.' then delete;
run;
/*DATA STEP THAT ONLY INCLUDES GENES FOUND IN ALL FIVE
LISTS(TIMEPOINTS)*/
data genecluster5;
set merge_cluster;
count=sum(one1,one2,one3,one4,one5);
if count =5 then output;
run;

proc sort data = genecluster5;
by gene;
run;
```

```
/*SIGNIFICANT GENES PRESENT IN TIME 1 AND TIME 5 ONLY/*/
data genecluster_t1_t5;
set merge_cluster;
if one1='.' then delete;
if one2='1' then delete;
if one3='1' then delete;
if one4='1' then delete;
if one5='.' then delete;
run;

proc sort data = genecluster_t1_t5;
by gene;
run;

/*SIGNIFICANT GENES PRESENT IN TIME 1 ONLY*/
data genecluster_t1;
set merge_cluster;
if one1='.' then delete;
if one2='1' then delete;
if one3='1' then delete;
if one4='1' then delete;
if one5='1' then delete;
run;

proc sort data = genecluster_t1;
by gene;
run;

/*SIGNIFICANT GENES PRESENT IN TIME 5 ONLY*/
data genecluster_t5;
set merge_cluster;
if one1='1' then delete;
if one2='1' then delete;
if one3='1' then delete;
if one4='1' then delete;
if one5='.' then delete;
run;

proc sort data = genecluster_t5;
by gene;
run;

/*READ IN ALL STANDARDIZED DATA*/
data povsntc;
infile 'C:\Carletha\povsntc.csv' dlm = "," dsd truncover lrecl=10000 firstobs=2;
```

```
input SubRow SubCol RowwiSub ColwiSub ControlAa1 TreatmentBa1 ControlAb1
TreatmentBb1 TreatmentAc1 ControlBc1 TreatmentAd1 ControlBd1 ControlAa2
TreatmentBa2 TreatmentAc2 ControlBc2 TreatmentAd2 ControlBd2 ControlAa3
TreatmentBa3 ControlAb3 TreatmentBb3 TreatmentAc3 ControlBc3 TreatmentAd3
ControlBd3 ControlAa4 TreatmentBa4 ControlAb4 TreatmentBb4 TreatmentAc4
ControlBc4 TreatmentAd4 ControlBd4 ControlAa5 TreatmentBa5 ControlAb5
TreatmentBb5 TreatmentAc5 ControlBc5 TreatmentAd5 ControlBd5;
run;

/*USED THIS EDITED UNIGENE TO CREATE AN ANNOTATION VARIABLE
FOR GENENETWORK AND TIGHTCLUST*/
data unigene_a;
infile 'C:\Carletha\unigene_a.csv' dsd truncover lrecl=10000 firstobs=2;
input SubRow SubCol RowwiSub ColwiSub Gene $ Gi : 10. TUC_ID $ : 30. annotation
$ : 80. Homo1_Desc $ : 80. Homo1_Species $ : 80. Homo2_PID $ : 80.
Homo2_Desc $ : 80. Homo2_Species $ : 80. Homo3_PID $ : 80. Homo3_Desc $ : 80.
Homo3_Species $ : 80.;
drop Gi TUC_ID;
if gene='BLANK' then delete;
if gene='blank' then delete;
if gene='xxxx' then delete;
if gene='Blank' then delete;
run;

/*TOOK THE MEAN VALUES OF THE DUPLICATES*/
data t1_control;
set povsntc;
keep SubRow SubCol RowwiSub ColwiSub ControlAa1 ControlAb1 ControlBc1
ControlBd1 t1con;
t1con=mean(ControlAa1,ControlAb1,ControlBc1,ControlBd1);
run;

proc sort data = t1_control;
by SubRow SubCol RowwiSub ColwiSub;
run;

data t1_treatment;
set povsntc;
keep SubRow SubCol RowwiSub ColwiSub TreatmentBa1 TreatmentBb1 TreatmentAc1
TreatmentAd1 t1tre;
t1tre=mean(TreatmentBa1, TreatmentBb1, TreatmentAc1, TreatmentAd1);
run;

proc sort data = t1_treatment;
by SubRow SubCol RowwiSub ColwiSub;
run;
```

```
data t2_control;
set povsntc;
keep SubRow SubCol RowwiSub ColwiSub ControlAa2 ControlBc2 ControlBd2 t2con;
t2con=mean(ControlAa2,ControlBc2, ControlBd2);
run;

proc sort data = t2_control;
by SubRow SubCol RowwiSub ColwiSub;
run;

data t2_treatment;
set povsntc;
keep SubRow SubCol RowwiSub ColwiSub TreatmentBa2 TreatmentAc2 TreatmentAd2
t2tre;
t2tre=mean(TreatmentBa2,TreatmentAc2,TreatmentAd2);
run;

proc sort data = t2_treatment;
by SubRow SubCol RowwiSub ColwiSub;
run;

data t3_control;
set povsntc;
keep SubRow SubCol RowwiSub ColwiSub ControlAa3 ControlAb3 ControlBc3
ControlBd3 t3con;
t3con=mean(ControlAa3, ControlAb3, ControlBc3, ControlBd3);
run;

proc sort data = t3_control;
by SubRow SubCol RowwiSub ColwiSub;
run;

data t3_treatment;
set povsntc;
keep SubRow SubCol RowwiSub ColwiSub TreatmentBa3 TreatmentBb3 TreatmentAc3
TreatmentAd3 t3tre;
t3tre=mean(TreatmentBa3, TreatmentBb3, TreatmentAc3, TreatmentAd3);
run;

proc sort data = t3_treatment;
by SubRow SubCol RowwiSub ColwiSub;
run;

data t4_control;
set povsntc;
```

```
keep SubRow SubCol RowwiSub ColwiSub ControlAa4 ControlAb4 ControlBc4
ControlBd4 t4con;
t4con=mean(ControlAa4, ControlAb4, ControlBc4, ControlBd4);
run;

proc sort data = t4_control;
by SubRow SubCol RowwiSub ColwiSub;
run;

data t4_treatment;
set povsntc;
keep SubRow SubCol RowwiSub ColwiSub TreatmentBa4 TreatmentBb4 TreatmentAc4
TreatmentAd4 t4tre;
t4tre=mean(TreatmentBa4, TreatmentBb4, TreatmentAc4, TreatmentAd4);
run;

proc sort data = t4_treatment;
by SubRow SubCol RowwiSub ColwiSub;
run;

data t5_control;
set povsntc;
keep SubRow SubCol RowwiSub ColwiSub ControlAa5 ControlAb5 ControlBc5
ControlBd5 t5con;
t5con=mean(ControlAa5, ControlAb5, ControlBc5, ControlBd5);
run;

proc sort data = t5_control;
by SubRow SubCol RowwiSub ColwiSub;
run;

data t5_treatment;
set povsntc;
keep SubRow SubCol RowwiSub ColwiSub TreatmentBa5 TreatmentBb5 TreatmentAc5
TreatmentAd5 t5tre;
t5tre=mean(TreatmentBa5, TreatmentBb5, TreatmentAc5, TreatmentAd5);
run;

proc sort data = t5_treatment;
by SubRow SubCol RowwiSub ColwiSub;
run;

/*MERGED ALL MEAN (FOR DUPLICATES WITHIN AN ARRAY)
STANDARDIZED VALUES TOGETHER*/
data time_means;
```

```
merge t1_control t1_treatment t2_control t2_treatment t3_control t3_treatment t4_control
t4_treatment t5_control t5_treatment;
by SubRow SubCol RowwiSub ColwiSub;
keep  SubRow SubCol RowwiSub ColwiSub t1con t1tre t2con t2tre t3con t3tre t4con
t4tre t5con t5tre;
run;

proc sort data = time_means;
by SubRow SubCol RowwiSub ColwiSub;
run;
proc sort data = unigene;
by SubRow SubCol RowwiSub ColwiSub;
run;

/*MERGED MEAN VALUES WITH GENE INFORMATION*/
data means_unigene;
merge time_means unigene;
by SubRow SubCol RowwiSub ColwiSub;
if t1con-t5tre = '.' then delete;
if gene =' ' then delete;
keep Gene t1con t1tre t2con t2tre t3con t3tre t4con t4tre t5con t5tre;
run;

proc sort data = means_unigene;
by gene;
run;

/*CREATED TO SHOW WHICH GENES WERE FOUND IN ONE SPECIFIC
TIMEPOINT ONLY FOR A 1% FDR*/
proc sort data=unigene;
by SubRow SubCol RowwiSub ColwiSub;
run;
data means_unigene_only;
merge time_means unigene;
by SubRow SubCol RowwiSub ColwiSub;
if t1con-t5tre = '.' then delete;
if gene =' ' then delete;
run;
proc sort data=means_unigene_only;
by gene;
run;

data t1only;
set genecluster_t1;
merge genecluster_t1 means_unigene_only;
by gene;
```

```
drop SubRow SubCol RowwiSub ColwiSub pvalue1 one1 pvalue2 one2 pvalue3 one3
pvalue4 one4 pvalue5 one5 t1con t1tre t2con t2tre t3con t3tre t4con t4tre t5con t5tre;
run;
data t2only;
set merge_cluster;
if one1='1' then delete;
if one2='.' then delete;
if one3='1' then delete;
if one4='1' then delete;
if one5='1' then delete;
merge merge_cluster means_unigene_only;
by gene;
drop SubRow SubCol RowwiSub ColwiSub pvalue1 one1 pvalue2 one2 pvalue3 one3
pvalue4 one4 pvalue5 one5 t1con t1tre t2con t2tre t3con t3tre t4con t4tre t5con t5tre;
run;
data t3only;
set merge_cluster;
if one1='1' then delete;
if one2='1' then delete;
if one3='.' then delete;
if one4='1' then delete;
if one5='1' then delete;
merge merge_cluster means_unigene_only;
by gene;
drop SubRow SubCol RowwiSub ColwiSub pvalue1 one1 pvalue2 one2 pvalue3 one3
pvalue4 one4 pvalue5 one5 t1con t1tre t2con t2tre t3con t3tre t4con t4tre t5con t5tre;
run;
data t4only;
set merge_cluster;
if one1='1' then delete;
if one2='1' then delete;
if one3='1' then delete;
if one4='.' then delete;
if one5='1' then delete;
merge merge_cluster means_unigene_only;
by gene;
drop SubRow SubCol RowwiSub ColwiSub pvalue1 one1 pvalue2 one2 pvalue3 one3
pvalue4 one4 pvalue5 one5 t1con t1tre t2con t2tre t3con t3tre t4con t4tre t5con t5tre;
run;
data t5only;
set genecluster_t5;
merge genecluster_t5 means_unigene_only;
by gene;
drop SubRow SubCol RowwiSub ColwiSub pvalue1 one1 pvalue2 one2 pvalue3 one3
pvalue4 one4 pvalue5 one5 t1con t1tre t2con t2tre t3con t3tre t4con t4tre t5con t5tre;
run;
```

```
/*FORMATTED FOR ORIOGEN*/
data ORIOGEN_control;
set povsntc;
keep SubRow SubCol RowwiSub ColwiSub ControlAa1 ControlAb1 ControlBc1
ControlBd1 ControlAa2 ControlBc2 ControlBd2 ControlAa3 ControlAb3 ControlBc3
ControlBd3 ControlAa4 ControlAb4 ControlBc4 ControlBd4 ControlAa5 ControlAb5
ControlBc5 ControlBd5;
merge povsntc unigene;
by SubRow SubCol RowwiSub ColwiSub;
keep  Gene ControlAa1 ControlAb1 ControlBc1 ControlBd1 ControlAa2 ControlBc2
ControlBd2 ControlAa3 ControlAb3 ControlBc3 ControlBd3 ControlAa4 ControlAb4
ControlBc4 ControlBd4 ControlAa5 ControlAb5 ControlBc5 ControlBd5;
if ControlAa1-ControlBd5="." then delete;
if gene =' ' then delete;
run;

data ORIOGEN_treatment;
set povsntc;
keep SubRow SubCol RowwiSub ColwiSub TreatmentBa1 TreatmentBb1 TreatmentAc1
TreatmentAd1 TreatmentBa2 TreatmentAc2 TreatmentAd2 TreatmentBa3 TreatmentBb3
TreatmentAc3 TreatmentAd3 TreatmentBa4 TreatmentBb4 TreatmentAc4 TreatmentAd4
TreatmentBa5 TreatmentBb5 TreatmentAc5 TreatmentAd5;
merge povsntc unigene;
by SubRow SubCol RowwiSub ColwiSub;
keep  Gene TreatmentBa1 TreatmentBb1 TreatmentAc1 TreatmentAd1 TreatmentBa2
TreatmentAc2 TreatmentAd2 TreatmentBa3 TreatmentBb3 TreatmentAc3 TreatmentAd3
TreatmentBa4 TreatmentBb4 TreatmentAc4 TreatmentAd4 TreatmentBa5 TreatmentBb5
TreatmentAc5 TreatmentAd5;
if TreatmentBa1-TreatmentAd5="." then delete;
if gene =' ' then delete;
run;

data mtable;
infile 'C:\Carletha\mtable.csv' dlm="," dsd truncover lrecl=1000 firstobs=2;
input SubRow SubCol RowwiSub ColwiSub m1  m2  m3a  m4a  m5  m6  m7a  m8a
m9  m10  m11a  m12a  m13  m14  m15a  m16a  m17  m18  m19a  m20a ;
run;

proc sort data=mtable;
by SubRow SubCol RowwiSub ColwiSub;
run;

proc sort data=unigene;
by SubRow SubCol RowwiSub ColwiSub;
run;
```

```
data ORIOGEN_m;
merge mtable unigene;
by SubRow SubCol RowwiSub ColwiSub;
if m1-m20a ='.' then delete;
if gene = ' ' then delete;
keep m1 m2 m3a m4a m5 m7a m8a m9 m10 m11a m12a m13 m14 m15a m16a m17 m18
m19a m20a gene;
run;


/*DETERMINED HOW MANY GENES IN THE M PROFILES ARE FOUND IN
EITHER THE TRT OR CON PROFILE LISTS*/
data check;
merge con_geneprofile trt_geneprofile m_geneprofile;
by gene;
if profileM ='.' then delete;
if profileC='.' AND profileT='.' then delete;
run;


/*IMPORTED GO_ID FILES FROM TRISTIAN*/
data outi_goid;
length goid1-goid87 $10;
infile 'C:\Carletha\GOID analysis files\outi_goid.csv' dsd truncover lrecl=10000
firstobs=2;
input gene $ tc $ goid1-goid87 ;
run;

data outi2_goid;
length goid1-goid85 $10;
infile 'C:\Carletha\GOID analysis files\outi2_goid.csv' dsd truncover lrecl=10000
firstobs=2;
input gene $ tc $ goid1-goid85;
run;

data all_goid;
set outi_goid outi2_goid;
run;

/*MERGED ONTOLOGY FILE WITH FINAL PROFILES*/
data con_profiles;
infile 'C:\Carletha\Profile results (final)\final_controlprofiles.csv' dlm="," dsd truncover
lrecl=10000 firstobs =2;
input gene $ profileC;
proc sort data=con_profiles;
```

```
by gene;
run;

proc sort data=all_goid;
by gene;
run;

data final_con_GO_profile;
merge all_goid con_profiles;
by gene;
if profileC =" " then delete;
run;
proc sort data=final_con_GO_profile;
by profileC;
run;

data trt_profiles;
infile 'C:\Carletha\Profile results (final)\final_trtprofiles.csv' dsd truncover lrecl=10000
firstobs =2;
input gene $ profileT;
proc sort data=trt_profiles;
by gene;
run;

data final_trt_GO_profile;
merge all_goid trt_profiles;
by gene;
if profileT =" " then delete;
proc sort data=final_trt_GO_profile;
by profileT;
run;

/*MERGED ORIGINAL HOMOLOG TEXT WITH ONLY LISTS, AND TRT
PROFILE LISTS*/
data t1only;
infile 'C:\Carletha\Significant gene lists (final)\t1only.csv' dsd truncover lrecl=10000
firstobs=2;
input gene $ Homo1_PID $ Homo1_Desc $ Homo1_Species $ Homo2_PID $
Homo2_Desc $ Homo2_Species $ Homo3_PID $ Homo3_Desc $ Homo3_Species $
count;
count=1;
keep gene count;
run;

data t2only;
```

```
infile 'C:\Carletha\Significant gene lists (final)\t2only.csv' dsd truncover lrecl=10000
firstobs=2;
input gene $ Homo1_PID $ Homo1_Desc $ Homo1_Species $ Homo2_PID $
Homo2_Desc $ Homo2_Species $ Homo3_PID $ Homo3_Desc $ Homo3_Species $
count;
count=2;
keep gene count;
run;

data t3only;
infile 'C:\Carletha\Significant gene lists (final)\t3only.csv' dsd truncover lrecl=10000
firstobs=2;
input gene $ Homo1_PID $ Homo1_Desc $ Homo1_Species $ Homo2_PID $
Homo2_Desc $ Homo2_Species $ Homo3_PID $ Homo3_Desc $ Homo3_Species $
count;
count=3;
keep gene count;
run;

data t4only;
infile 'C:\Carletha\Significant gene lists (final)\t4only.csv' dsd truncover lrecl=10000
firstobs=2;
input gene $ Homo1_PID $ Homo1_Desc $ Homo1_Species $ Homo2_PID $
Homo2_Desc $ Homo2_Species $ Homo3_PID $ Homo3_Desc $ Homo3_Species $
count;
count=4;
keep gene count;
run;

data t5only;
infile 'C:\Carletha\Significant gene lists (final)\t5only.csv' dsd truncover lrecl=10000
firstobs=2;
input gene $ Homo1_PID $ Homo1_Desc $ Homo1_Species $ Homo2_PID $
Homo2_Desc $ Homo2_Species $ Homo3_PID $ Homo3_Desc $ Homo3_Species $
count;
count=5;
keep gene count;
run;

proc sort data=unigene;
by gene;
run;
proc sort data=t1only;
by gene;
run;
proc sort data=t2only;
```

```
by gene;
run;
proc sort data=t3only;
by gene;
run;
proc sort data=t4only;
by gene;
run;
proc sort data=t5only;
by gene;
run;

data t1only_homolog;
merge unigene t1only;
by gene;
if count='.' then delete;
if Homo1_Desc =' ' then delete;
keep Homo1_Desc count;
run;

data t2only_homolog;
merge unigene t2only;
by gene;
if count='.' then delete;
if Homo1_Desc =' ' then delete;
keep Homo1_Desc count;
run;

data t3only_homolog;
merge unigene t3only;
by gene;
if count='.' then delete;
if Homo1_Desc =' ' then delete;
keep Homo1_Desc count;
run;

data t4only_homolog;
merge unigene t4only;
by gene;
if count='.' then delete;
if Homo1_Desc =' ' then delete;
keep Homo1_Desc count;
run;

data t5only_homolog;
merge unigene t5only;
```

```
by gene;
if count='.' then delete;
if Homo1_Desc =' ' then delete;
keep Homo1_Desc count;
run;

data profile_lists;
infile 'C:\Carletha\Profile Results (final)\FINAL_trtprofiles.csv' dsd truncover firstobs=2;
input gene $ profile $;
run;

data trtpr1;
set profile_lists;
if profile=1 then output;
run;

data trtpr2;
set profile_lists;
if profile=2 then output;
run;

data trtpr3;
set profile_lists;
if profile=3 then output;
run;

data trtpr4;
set profile_lists;
if profile=4 then output;
run;

data trtpr5;
set profile_lists;
if profile=5 then output;
run;

data trtpr6;
set profile_lists;
if profile=6 then output;
run;

data trtpr7;
set profile_lists;
if profile=7 then output;
run;
```

```
data trtpr8;
set profile_lists;
if profile=8 then output;
run;

data trtpr1_homolog;
merge trtpr1 unigene;
by gene;
if profile=' ' then delete;
if Homo1_Desc =' ' then delete;
keep gene Homo1_Desc profile;
run;

data trtpr2_homolog;
merge trtpr2 unigene;
by gene;
if profile=' ' then delete;
if Homo1_Desc =' ' then delete;
keep gene Homo1_Desc profile;
run;

data trtpr3_homolog;
merge trtpr3 unigene;
by gene;
if profile=' ' then delete;
if Homo1_Desc =' ' then delete;
keep gene Homo1_Desc profile;
run;

data trtpr4_homolog;
merge trtpr4 unigene;
by gene;
if profile=' ' then delete;
if Homo1_Desc =' ' then delete;
keep gene Homo1_Desc profile;
run;

data trtpr5_homolog;
merge trtpr5 unigene;
by gene;
if profile=' ' then delete;
if Homo1_Desc =' ' then delete;
keep gene Homo1_Desc profile;
run;

data trtpr6_homolog;
```

```
merge trtpr6 unigene;
by gene;
if profile=' ' then delete;
if Homo1_Desc =' ' then delete;
keep gene Homo1_Desc profile;
run;

data trtpr7_homolog;
merge trtpr7 unigene;
by gene;
if profile=' ' then delete;
if Homo1_Desc =' ' then delete;
keep gene Homo1_Desc profile;
run;

data trtpr8_homolog;
merge trtpr8 unigene;
by gene;
if profile=' ' then delete;
if Homo1_Desc =' ' then delete;
keep gene Homo1_Desc profile;
run;

/*MERGED ONTOLOGY WITH SIGNIFICANT GENE LISTS*/
data t1;
infile 'C:\Carletha\Significant gene lists (final)\t1only.csv' dsd truncover lrecl=10000
firstobs=2;
input gene $ Homo1_PID $ Homo1_Desc $ Homo1_Species $ Homo2_PID $
Homo2_Desc $ Homo2_Species $ Homo3_PID $ Homo3_Desc $ Homo3_Species $
count;
keep gene count;
run;

data t1_GO;
merge t1 all_goid;
by gene;
drop tc;
if count=' ' then delete;
drop count;
run;

data t2;
infile 'C:\Carletha\Significant gene lists (final)\t2only.csv' dsd truncover lrecl=10000
firstobs=2;
```

```
input gene $ Homo1_PID $ Homo1_Desc $ Homo1_Species $ Homo2_PID $
Homo2_Desc $ Homo2_Species $ Homo3_PID $ Homo3_Desc $ Homo3_Species $
count;
keep gene count;
run;

data t2_GO;
merge t2 all_goid;
by gene;
drop tc;
if count=' ' then delete;
drop count;
run;

data t3;
infile 'C:\Carletha\Significant gene lists (final)\t3only.csv' dsd truncover lrecl=10000
firstobs=2;
input gene $ Homo1_PID $ Homo1_Desc $ Homo1_Species $ Homo2_PID $
Homo2_Desc $ Homo2_Species $ Homo3_PID $ Homo3_Desc $ Homo3_Species $
count;
keep gene count;
run;

data t3_GO;
merge t3 all_goid;
by gene;
drop tc;
if count=' ' then delete;
drop count;
run;

data t4;
infile 'C:\Carletha\Significant gene lists (final)\t4only.csv' dsd truncover lrecl=10000
firstobs=2;
input gene $ Homo1_PID $ Homo1_Desc $ Homo1_Species $ Homo2_PID $
Homo2_Desc $ Homo2_Species $ Homo3_PID $ Homo3_Desc $ Homo3_Species $
count;
keep gene count;
run;

data t4_GO;
merge t4 all_goid;
by gene;
drop tc;
if count=' ' then delete;
drop count;
```

```
run;

data t5;
infile 'C:\Carletha\Significant gene lists (final)\t5only.csv' dsd truncover lrecl=10000
firstobs=2;
input gene $ Homo1_PID $ Homo1_Desc $ Homo1_Species $ Homo2_PID $
Homo2_Desc $ Homo2_Species $ Homo3_PID $ Homo3_Desc $ Homo3_Species $
count;
keep gene count;
run;

data t5_GO;
merge t5 all_goid;
by gene;
drop tc;
if count=' ' then delete;
drop count;
run;

data t1_t5;
infile 'C:\Carletha\Significant gene lists (final)\t1_t5only.csv' dsd truncover lrecl=10000
firstobs=2;
input pvalue1 gene $ one1 pvalue2 one2 pvalue3 one3 pvalue4 one4 pvalue5 one5;
keep gene one1;
proc sort data=t1_t5;
by gene;
run;

data t1_t5GO;
merge t1_t5 all_goid;
by gene;
drop tc;
if one1=' ' then delete;
drop one1;
run;

/*READ IN SIGNIFICANT GENE LISTS FOR CHI-SQUARE ANALYSIS*/
data percent_allgoid;
length goid $10;
infile 'C:\Carletha\GOID analysis files\percent_allgoid.csv' dsd truncover lrecl=10000
firstobs=2;
input goid $ freq percent;
run;

data percent_t1goid;
length goid $10;
```

```
infile 'C:\Carletha\GOID analysis files\percent_t1goid.csv' dsd truncover lrecl=10000
firstobs=2;
input goid $ freq1 percent1;
run;

data percent_t2goid;
length goid $10;
infile 'C:\Carletha\GOID analysis files\percent_t2goid.csv' dsd truncover lrecl=10000
firstobs=2;
input goid $ freq2 percent2;
run;

data percent_t3goid;
length goid $10;
infile 'C:\Carletha\GOID analysis files\percent_t3goid.csv' dsd truncover lrecl=10000
firstobs=2;
input goid $ freq3 percent3;
run;

data percent_t4goid;
length goid $10;
infile 'C:\Carletha\GOID analysis files\percent_t4goid.csv' dsd truncover lrecl=10000
firstobs=2;
input goid $ freq4 percent4;
run;

data percent_t5goid;
length goid $10;
infile 'C:\Carletha\GOID analysis files\percent_t5goid.csv' dsd truncover lrecl=10000
firstobs=2;
input goid $ freq5 percent5;
run;

data percent_t1_t5goid;
length goid $10;
infile 'C:\Carletha\GOID analysis files\percent_t1_t5goid.csv' dsd truncover lrecl=10000
firstobs=2;
input goid $ freq15 percent15;
run;

/*MERGED MASTER FREQ WITH EACH LIST FREQ FOR CHI-SQUARE
ANALYSIS*/
data chisquare_t1;
merge percent_allgoid percent_t1goid;
by goid;
if percent1='.' then delete;
```

```
run;

data chisquare_t2;
merge percent_allgoid percent_t2goid;
by goid;
if percent2='.' then delete;
run;

data chisquare_t3;
merge percent_allgoid percent_t3goid;
by goid;
if percent3='.' then delete;
run;

data chisquare_t4;
merge percent_allgoid percent_t4goid;
by goid;
if percent4='.' then delete;
run;

data chisquare_t5;
merge percent_allgoid percent_t5goid;
by goid;
if percent5='.' then delete;
run;

data chisquare_t1_t5;
merge percent_allgoid percent_t1_t5goid;
by goid;
if percent15='.' then delete;
run;

/*READ IN FILE WITH 1% FDR FOR SIG OVER/UNDER REPRESENTED GOIDS
FOR EACH SIGNIFICANT LIST*/
data t1_over_under_go;
length goid $10;
infile 'C:\Carletha\Chi-square data\go_qvalues_t1.csv' dsd truncover lrecl=1000
firstobs=2;
input pvalue qvalue fdr1 fdr5 goid $;
if fdr1='1' then output;
keep fdr1 goid;
run;

data t2_over_under_go;
length goid $10;
```

```
infile 'C:\Carletha\Chi-square data\go_qvalues_t2.csv' dsd truncover lrecl=1000
firstobs=2;
input pvalue qvalue fdr1 fdr5 goid $;
if fdr1='1' then output;
keep fdr1 goid;
run;

data t3_over_under_go;
length goid $10;
infile 'C:\Carletha\Chi-square data\go_qvalues_t3.csv' dsd truncover lrecl=1000
firstobs=2;
input pvalue qvalue $ fdr1 fdr5 goid $;
if fdr1='1' then output;
keep fdr1 goid;
run;

data t4_over_under_go;
length goid $10;
infile 'C:\Carletha\Chi-square data\go_qvalues_t4.csv' dsd truncover lrecl=1000
firstobs=2;
input pvalue qvalue $ fdr1 fdr5 goid $;
if fdr1='1' then output;
keep fdr1 goid;
run;

data t5_over_under_go;
length goid $10;
infile 'C:\Carletha\Chi-square data\go_qvalues_t5.csv' dsd truncover lrecl=1000
firstobs=2;
input pvalue qvalue $ fdr1 fdr5 goid $;
if fdr1='1' then output;
keep fdr1 goid;
run;

data t1_t5over_under_go;
length goid $10;
infile 'C:\Carletha\Chi-square data\go_qvalues_t1t5.csv' dsd truncover lrecl=1000
firstobs=2;
input pvalue qvalue $ fdr1 fdr5 goid $;
if fdr1='1' then output;
keep fdr1 goid;
run;

/*READ IN CONTROL PROFILES FOR CHI-SQUARE ANALYSIS*/
data percent_allgoid;
length goid $10;
```

```
infile 'C:\Carletha\GOID analysis files\percent_allgoid.csv' dsd truncover lrecl=10000
firstobs=2;
input goid $ freq percent;
run;

data percent_conprofile1_go;
length goid $10;
infile 'C:\Carletha\GOID analysis files\percent_conprofile1_go.csv' dsd truncover
lrecl=10000 firstobs=2;
input goid $ freq1 percent1;
run;

data percent_conprofile2_go;
length goid $10;
infile 'C:\Carletha\GOID analysis files\percent_conprofile2_go.csv' dsd truncover
lrecl=10000 firstobs=2;
input goid $ freq2 percent2;
run;

data percent_conprofile3_go;
length goid $10;
infile 'C:\Carletha\GOID analysis files\percent_conprofile3_go.csv' dsd truncover
lrecl=10000 firstobs=2;
input goid $ freq3 percent3;
run;

data percent_conprofile4_go;
length goid $10;
infile 'C:\Carletha\GOID analysis files\percent_conprofile4_go.csv' dsd truncover
lrecl=10000 firstobs=2;
input goid $ freq4 percent4;
run;

data percent_conprofile5_go;
length goid $10;
infile 'C:\Carletha\GOID analysis files\percent_conprofile5_go.csv' dsd truncover
lrecl=10000 firstobs=2;
input goid $ freq5 percent5;
run;

data percent_conprofile6_go;
length goid $10;
infile 'C:\Carletha\GOID analysis files\percent_conprofile6_go.csv' dsd truncover
lrecl=10000 firstobs=2;
input goid $ freq6 percent6;
run;
```

```
data percent_conprofile7_go;
length goid $10;
infile 'C:\Carletha\GOID analysis files\percent_conprofile7_go.csv' dsd truncover
lrecl=10000 firstobs=2;
input goid $ freq7 percent7;
run;

data percent_conprofile8_go;
length goid $10;
infile 'C:\Carletha\GOID analysis files\percent_conprofile8_go.csv' dsd truncover
lrecl=10000 firstobs=2;
input goid $ freq8 percent8;
run;

/*READ IN TREATMENT PROFILES FOR CHI-SQUARE ANALYSIS*/
data percent_trtprofile1_go;
length goid $10;
infile 'C:\Carletha\GOID analysis files\percent_trtprofile1_go.csv' dsd truncover
lrecl=10000 firstobs=2;
input goid $ freq1 percent1;
run;

data percent_trtprofile2_go;
length goid $10;
infile 'C:\Carletha\GOID analysis files\percent_trtprofile2_go.csv' dsd truncover
lrecl=10000 firstobs=2;
input goid $ freq2 percent2;
run;

data percent_trtprofile3_go;
length goid $10;
infile 'C:\Carletha\GOID analysis files\percent_trtprofile3_go.csv' dsd truncover
lrecl=10000 firstobs=2;
input goid $ freq3 percent3;
run;

data percent_trtprofile4_go;
length goid $10;
infile 'C:\Carletha\GOID analysis files\percent_trtprofile4_go.csv' dsd truncover
lrecl=10000 firstobs=2;
input goid $ freq4 percent4;
run;

data percent_trtprofile5_go;
length goid $10;
```

```
infile 'C:\Carletha\GOID analysis files\percent_trtprofile5_go.csv' dsd truncover
lrecl=10000 firstobs=2;
input goid $ freq5 percent5;
run;

data percent_trtprofile6_go;
length goid $10;
infile 'C:\Carletha\GOID analysis files\percent_trtprofile6_go.csv' dsd truncover
lrecl=10000 firstobs=2;
input goid $ freq6 percent6;
run;

data percent_trtprofile7_go;
length goid $10;
infile 'C:\Carletha\GOID analysis files\percent_trtprofile7_go.csv' dsd truncover
lrecl=10000 firstobs=2;
input goid $ freq7 percent7;
run;

data percent_trtprofile8_go;
length goid $10;
infile 'C:\Carletha\GOID analysis files\percent_trtprofile8_go.csv' dsd truncover
lrecl=10000 firstobs=2;
input goid $ freq8 percent8;
run;

/*MERGED MASTER FREQ WITH EACH PROFILE LIST FREQ FOR CHI-SQUARE
ANALYSIS FOR CONTROL PROFILES*/
data chisquare_p1_con;
merge percent_allgoid percent_conprofile1_go;
by goid;
if percent1='.' then delete;
run;

data chisquare_p2_con;
merge percent_allgoid percent_conprofile2_go;
by goid;
if percent2='.' then delete;
run;

data chisquare_p3_con;
merge percent_allgoid percent_conprofile3_go;
by goid;
if percent3='.' then delete;
run;
```

```
data chisquare_p4_con;
merge percent_allgoid percent_conprofile4_go;
by goid;
if percent4='.' then delete;
run;

data chisquare_p5_con;
merge percent_allgoid percent_conprofile5_go;
by goid;
if percent5='.' then delete;
run;

data chisquare_p6_con;
merge percent_allgoid percent_conprofile6_go;
by goid;
if percent6='.' then delete;
run;

data chisquare_p7_con;
merge percent_allgoid percent_conprofile7_go;
by goid;
if percent7='.' then delete;
run;

data chisquare_p8_con;
merge percent_allgoid percent_conprofile8_go;
by goid;
if percent8='.' then delete;
run;


/*MERGED MASTER FREQ WITH EACH PROFILE LIST FREQ FOR CHI-SQUARE
ANALYSIS FOR TREATMENT PROFILES*/
data chisquare_p1_trt;
merge percent_allgoid percent_trtprofile1_go;
by goid;
if percent1='.' then delete;
run;

data chisquare_p2_trt;
merge percent_allgoid percent_trtprofile2_go;
by goid;
if percent2='.' then delete;
run;

data chisquare_p3_trt;
```

```
merge percent_allgoid percent_trtprofile3_go;
by goid;
if percent3='.' then delete;
run;

data chisquare_p4_trt;
merge percent_allgoid percent_trtprofile4_go;
by goid;
if percent4='.' then delete;
run;

data chisquare_p5_trt;
merge percent_allgoid percent_trtprofile5_go;
by goid;
if percent5='.' then delete;
run;

data chisquare_p6_trt;
merge percent_allgoid percent_trtprofile6_go;
by goid;
if percent6='.' then delete;
run;

data chisquare_p7_trt;
merge percent_allgoid percent_trtprofile7_go;
by goid;
if percent7='.' then delete;
run;

data chisquare_p8_trt;
merge percent_allgoid percent_trtprofile8_go;
by goid;
if percent8='.' then delete;
run;


/*READ IN FILE WITH 1% FDR FOR SIG OVER/UNDER REPRESENTED GOIDS
FOR EACH CONTROL PROFILE FROM Q RESULTS*/
data go_qvalues_con_profile1;
length goid $10;
infile 'C:\Carletha\Chi-square data\qvalues_con_p1.csv' dsd truncover lrecl=1000
firstobs=2;
input pvalue qvalue fdr1 goid $;
if fdr1='1' then output;
keep fdr1 goid;
run;
```

```
data go_qvalues_con_profile2;
length goid $10;
infile 'C:\Carletha\Chi-square data\qvalues_con_p2.csv' dsd truncover lrecl=1000
firstobs=2;
input pvalue qvalue fdr1 goid $;
if fdr1='1' then output;
keep fdr1 goid;
run;

data go_qvalues_con_profile3;
length goid $10;
infile 'C:\Carletha\Chi-square data\qvalues_con_p3.csv' dsd truncover lrecl=1000
firstobs=2;
input pvalue qvalue fdr1 goid $;
if fdr1='1' then output;
keep fdr1 goid;
run;

data go_qvalues_con_profile4;
length goid $10;
infile 'C:\Carletha\Chi-square data\qvalues_con_p4.csv' dsd truncover lrecl=1000
firstobs=2;
input pvalue qvalue fdr1 goid $;
if fdr1='1' then output;
keep fdr1 goid;
run;

data go_qvalues_con_profile5;
length goid $10;
infile 'C:\Carletha\Chi-square data\qvalues_con_p5.csv' dsd truncover lrecl=1000
firstobs=2;
input pvalue qvalue fdr1 goid $;
if fdr1='1' then output;
keep fdr1 goid;
run;

data go_qvalues_con_profile6;
length goid $10;
infile 'C:\Carletha\Chi-square data\qvalues_con_p6.csv' dsd truncover lrecl=1000
firstobs=2;
input pvalue qvalue fdr1 goid $;
if fdr1='1' then output;
keep fdr1 goid;
run;
```

```
data go_qvalues_con_profile7;
length goid $10;
infile 'C:\Carletha\Chi-square data\qvalues_con_p7.csv' dsd truncover lrecl=1000
firstobs=2;
input pvalue qvalue fdr1 goid $;
if fdr1='1' then output;
keep fdr1 goid;
run;

data go_qvalues_con_profile8;
length goid $10;
infile 'C:\Carletha\Chi-square data\qvalues_con_p8.csv' dsd truncover lrecl=1000
firstobs=2;
input pvalue qvalue fdr1 goid $;
if fdr1='1' then output;
keep fdr1 goid;
run;


/*CREATED FILE WITH 1% FDR FOR SIG OVER/UNDER REPRESENTED GOIDS
FOR EACH TRT PROFILE FROM Q RESULTS*/
data go_qvalues_trt_profile1;
length goid $10;
infile 'C:\Carletha\Chi-square data\qvalues_trt_p1.csv' dsd truncover lrecl=1000
firstobs=2;
input pvalue qvalue fdr1 goid $;
if fdr1='1' then output;
keep fdr1 goid;
run;

data go_qvalues_trt_profile2;
length goid $10;
infile 'C:\Carletha\Chi-square data\qvalues_trt_p2.csv' dsd truncover lrecl=1000
firstobs=2;
input pvalue qvalue fdr1 goid $;
if fdr1='1' then output;
keep fdr1 goid;
run;

data go_qvalues_trt_profile3;
length goid $10;
infile 'C:\Carletha\Chi-square data\qvalues_trt_p3.csv' dsd truncover lrecl=1000
firstobs=2;
input pvalue qvalue fdr1 goid $;
if fdr1='1' then output;
keep fdr1 goid;
```

```
run;

data go_qvalues_trt_profile4;
length goid $10;
infile 'C:\Carletha\Chi-square data\qvalues_trt_p4.csv' dsd truncover lrecl=1000
firstobs=2;
input pvalue qvalue fdr1 goid $;
if fdr1='1' then output;
keep fdr1 goid;
run;

data go_qvalues_trt_profile5;
length goid $10;
infile 'C:\Carletha\Chi-square data\qvalues_trt_p5.csv' dsd truncover lrecl=1000
firstobs=2;
input pvalue qvalue fdr1 goid $;
if fdr1='1' then output;
keep fdr1 goid;
run;

data go_qvalues_trt_profile6;
length goid $10;
infile 'C:\Carletha\Chi-square data\qvalues_trt_p6.csv' dsd truncover lrecl=1000
firstobs=2;
input pvalue qvalue fdr1 goid $;
if fdr1='1' then output;
keep fdr1 goid;
run;

data go_qvalues_trt_profile7;
length goid $10;
infile 'C:\Carletha\Chi-square data\qvalues_trt_p7.csv' dsd truncover lrecl=1000
firstobs=2;
input pvalue qvalue fdr1 goid $;
if fdr1='1' then output;
keep fdr1 goid;
run;

data go_qvalues_trt_profile8;
length goid $10;
infile 'C:\Carletha\Chi-square data\qvalues_trt_p8.csv' dsd truncover lrecl=1000
firstobs=2;
input pvalue qvalue fdr1 goid $;
if fdr1='1' then output;
keep fdr1 goid;
run;
```

```
/*DETERMINED WHICH SIG. OVER/UNDER GOID IS ONLY PRESENT IN
MASTER LIST ONCE, AND DELETED FROM RESULTS*/
data t1_goidcheck;
length goid $10;
infile 'C:\Carletha\OverUnder represented GOID lists\t1_over_under_go.csv' dsd
truncover lrecl=1000 firstobs=2;
input goid $ fdr1 text $: 10000.;
run;

data t3_goidcheck;
length goid $10;
infile 'C:\Carletha\OverUnder represented GOID lists\t3_over_under_go.csv' dsd
truncover lrecl=1000 firstobs=2;
input goid $ fdr1 text $: 10000.;
run;

data con_p1_check;
length goid $10;
infile 'C:\Carletha\OverUnder represented GOID lists\final_con_p1_go.csv' dsd truncover
lrecl=1000 firstobs=2;
input goid $ fdr1 text $: 10000.;
run;

data con_p2_check;
length goid $10;
infile 'C:\Carletha\OverUnder represented GOID lists\final_con_p2_go.csv' dsd truncover
lrecl=1000 firstobs=2;
input goid $ fdr1 text $: 10000.;
run;

data trt_p2_check;
length goid $10;
infile 'C:\Carletha\OverUnder represented GOID lists\final_trt_p2_go.csv' dsd truncover
lrecl=1000 firstobs=2;
input goid $ fdr1 text $: 10000.;
run;

data trt_p4_check;
length goid $10;
infile 'C:\Carletha\OverUnder represented GOID lists\final_trt_p4_go.csv' dsd truncover
lrecl=1000 firstobs=2;
input goid $ fdr1 text $: 10000.;
run;

data trt_p5_check;
```

```
length goid $10;
infile 'C:\Carletha\OverUnder represented GOID lists\final_trt_p5_go.csv' dsd truncover
lrecl=1000 firstobs=2;
input goid $ fdr1 text $: 10000.;
run;

data trt_p7_check;
length goid $10;
infile 'C:\Carletha\OverUnder represented GOID lists\final_trt_p7_go.csv' dsd truncover
lrecl=1000 firstobs=2;
input goid $ fdr1 text $: 10000.;
run;

data allgoid_check;
length goid $10;
infile 'C:\Carletha\GOID analysis files\percent_allgoid.csv' dsd truncover lrecl=1000
firstobs=2;
input goid $ freq percent ;
run;

data t1_allgoid;
merge allgoid_check t1_goidcheck;
by goid;
if fdr1='.' then delete;
keep goid freq text;
if freq='1' then delete;
run;

data t3_allgoid;
merge allgoid_check t3_goidcheck;
by goid;
if fdr1='.' then delete;
keep goid freq text;
if freq='1' then delete;
run;

data conp1_allgoid;
merge allgoid_check con_p1_check;
by goid;
if fdr1='.' then delete;
keep goid freq text;
if freq='1' then delete;
run;

data conp2_allgoid;
merge allgoid_check con_p2_check;
```

```
by goid;
if fdr1='.' then delete;
keep goid freq text;
if freq='1' then delete;
run;

data trtp2_allgoid;
merge allgoid_check trt_p2_check;
by goid;
if fdr1='.' then delete;
keep goid freq text;
if freq='1' then delete;
run;

data trtp4_allgoid;
merge allgoid_check trt_p4_check;
by goid;
if fdr1='.' then delete;
keep goid freq text;
if freq='1' then delete;
run;

data trtp5_allgoid;
merge allgoid_check trt_p5_check;
by goid;
if fdr1='.' then delete;
keep goid freq text;
if freq='1' then delete;
run;

data trtp7_allgoid;
merge allgoid_check trt_p7_check;
by goid;
if fdr1='.' then delete;
if freq='1' then delete;
run;

data edited_unigene;
set unigene;
drop SubRow SubCol RowwiSub ColwiSub Gi TUC_ID;
proc sort data=edited_unigene;
by gene;
run;
proc sort data=all_goid;
by gene;
run;
```

```
data allgoid_unigene;
merge all_goid edited_unigene;
by Gene;
run;
```

```
/*READ IN PROFILES TO OBTAIN GOID FOR GENES IN EACH PROFILE*/
data m_geneprofile;
infile 'C:\Carletha\Profile Results (final)\m_gene&profile.csv' dlm="," dsd truncover
lrecl=10000;
input gene $ profileM;
run;
```

```
data con_geneprofile;
infile 'C:\Carletha\Profile Results (final)\control_gene&profile.csv' dlm="," dsd truncover
lrecl=10000;
input gene $ profileC;
run;

data trt_geneprofile;
infile 'C:\Carletha\Profile Results (final)\treatment_gene&profile.csv' dlm="," dsd
truncover lrecl=10000;
input gene $ profileT;
run;

proc sort data=m_geneprofile;
by gene;
run;

proc sort data=con_geneprofile;
by gene;
run;

proc sort data=trt_geneprofile;
by gene;
run;

/*COMPARISON TO SEE WHICH GENES HAVE THE SAME PROFILES
BETWEEN TREATMENT AND CONTROL*/
data profile_comparison;
merge con_geneprofile trt_geneprofile;
by gene;
if profileC ='.' then delete;
if profileT ='.' then delete;
if profileC=profileT then output;
proc sort data=profile_comparison;
by profileC;
run;

data same_profiles;
infile 'C:\Carletha\profile_comparison.csv' dlm="," dsd truncover lrecl=10000 firstobs
=2;
input gene $ profileC profileT;
proc sort data=same_profiles;
by gene;
run;

proc sort data=con_geneprofile;
```

```
by gene;
run;

proc sort data=trt_geneprofile;
by gene;
run;

data final_controlprofiles;
merge con_geneprofile same_profiles;
if profileC=profileT then delete;
drop profileT;
proc sort data=final_controlprofiles;
by profileC;
run;

data final_trtprofiles;
merge trt_geneprofile same_profiles;
if profileC=profileT then delete;
drop profileC;
proc sort data=final_trtprofiles;
by profileT;
run;


/*READ IN GENES FOUND IN TREATMENT PROFILES FOR NETWORK
ANALYSIS*/
data trt_profileGN;
set trt_geneprofile;
run;

/*PREPARED TO RUN GENENETWORK FOR GENES FOUND IN TREATMENT
PROFILES 3 & 6*/
data trtprofile3_6;
set trt_profileGN;
if Profile='1' then delete;
if Profile='2' then delete;
if Profile='4' then delete;
if Profile='5' then delete;
if Profile='7' then delete;
if Profile='8' then delete;
run;

proc sort data=trtprofile3_6;
by gene;
run;
```

```
proc sort data=unigene_a;
by gene;
run;

data meansunigene_unigenea;
merge means_unigene unigene_a;
by gene;
keep gene annotation t1con t1tre t2con t2tre t3con t3tre t4con t4tre t5con t5tre;
if t1con-t5tre='.' then delete;
run;

data unigene_trtprofile3_6;
merge meansunigene_unigenea trtprofile3_6;
by gene;
if profile='.' then delete;
run;

proc sort data=unigene_trtprofile3_6;
by profile;
run;

data GNtreatment3_6;
set unigene_trtprofile3_6;
if profile=' ' then delete;
keep Gene annotation t1tre t2tre t3tre t4tre t5tre;
run;

/*READ IN GENES FOUND IN CONTROL PROFILES FOR NETWORK
ANALYSIS*/
data con_profileGN;
set con_geneprofile;
run;

/*PREPARED TO RUN GENENETWORK FOR GENES FOUND IN CONTROL
PROFILES 3 & 6*/
data conprofile3_6;
set con_profileGN;
if Profile='1' then delete;
if Profile='2' then delete;
if Profile='4' then delete;
if Profile='5' then delete;
if Profile='7' then delete;
if Profile='8' then delete;
run;

proc sort data=conprofile3_6;
```

```
by gene;
run;

proc sort data=unigene_a;
by gene;
run;

data meansunigene_unigenea;
merge means_unigene unigene_a;
by gene;
keep gene annotation t1con t1tre t2con t2tre t3con t3tre t4con t4tre t5con t5tre;
if t1con-t5tre='.' then delete;
run;

data unigene_conprofile3_6;
merge meansunigene_unigenea conprofile3_6;
by gene;
if profile='.' then delete;
run;

proc sort data=unigene_conprofile3_6;
by profile;
run;

data GNcontrol3_6;
set unigene_conprofile3_6;
keep Gene annotation t1con t2con t3con t4con t5con;
if profile=' ' then delete;
run;

/*DATA SETS CREATED FOR TIGHTCLUST*/
/*CREATED TO DETERMINE IF TIGHTCLUST PROVIDED SIMILIAR RESULTS
AS ORIOGEN FOR EACH PROFILE SET*/
data trt_profiles;
set trt_geneprofile;
run;

data cluster_trtprofiles;
merge trt_profiles means_unigene;
by gene;
if profileT='.' or profileT=' ' then delete;
drop profileT;
merge cluster_trtprofiles unigene_a;
keep gene t1con t1tre t2con t2tre t3con t3tre t4con t4tre t5con t5tre annotation;
run;
```

```
data con_profiles;
set con_geneprofile;
run;

data cluster_conprofiles;
merge con_profiles means_unigene;
by gene;
if profileC='.' or profileC=' ' then delete;
drop profileC;
merge cluster_conprofiles unigene_a;
keep gene t1con t1tre t2con t2tre t3con t3tre t4con t4tre t5con t5tre annotation;
run;

data m_profiles;
set m_geneprofile;
run;

data cluster_mprofiles;
merge m_profiles means_unigene;
by gene;
if profileM='.' or profileM=' ' then delete;
drop profileM;
merge cluster_mprofiles unigene_a;
keep gene t1con t1tre t2con t2tre t3con t3tre t4con t4tre t5con t5tre annotation;
run;
```

APPENDIX B

R code for vsn standardization

```
library (van)
outvsn<-vsn(timecourse)
write.table(exprs(outvsn),"posttimecourse.csv",sep=",")
printtip<-read.table("printtip.csv",sep=",",header=TRUE)
outvsn2<-vsn(timecourse,strate=printtip)
```

APPENDIX C

R code for MvA and dye duplicate scatterplots

```
#READING IN STANDARDIZED DATA
vsn1<-read.table("C:\\Carletha\\postvsntimecourse.csv",sep=",",header=FALSE)

#CHECKING TO MAKE SURE THE DATA WAS READ IN CORRECTLY
vsn1[1:5,5:6]

#CODING MVA PLOTS MANUALLY
g1stand<-vsn1[,5]
r1stand<-vsn1[,6]
m1<-r1stand-g1stand
a1<-(r1stand+g1stand)/2
plot(a1,m1, xlab="A1 (average of log pixel intensities)",ylab="M1 (difference in log
pixel intensity)", type='p',pch='.')

g2stand<-vsn1[,7]
r2stand<-vsn1[,8]
m2<-r2stand-g2stand
a2<-(r2stand+g2stand)/2
plot(a2,m2,xlab="A2 (average of log pixel intensities)",ylab="M2 (difference in log pixel
intensity)", type='p',pch='.')

g3stand<-vsn1[,9]
r3stand<-vsn1[,10]
m3<-r3stand-g3stand
a3<-(r3stand+g3stand)/2
plot(a3,m3,xlab="A3 (average of log pixel intensities)",ylab="M3 (difference in log pixel
intensity)", type='p',pch='.')


g4stand<-vsn1[,11]
r4stand<-vsn1[,12]
m4<-r4stand-g4stand
a4<-(r4stand+g4stand)/2
plot(a4,m4,xlab="A4 (average of log pixel intensities)",ylab="M4 (difference in log pixel
intensity)", type='p',pch='.')


g5stand<-vsn1[,13]
r5stand<-vsn1[,14]
m5<-r5stand-g5stand
a5<-(r5stand+g5stand)/2
```

```
plot(a5,m5,xlab="A5 (average of log pixel intensities)",ylab="M5 (difference in log pixel intensity)", type='p',pch='.')

g6stand<-vsn1[,15]
r6stand<-vsn1[,16]
m6<-r6stand-g6stand
a6<-(r6stand+g6stand)/2
plot(a6,m6,xlab="A6 (average of log pixel intensities)",ylab="M6 (difference in log pixel intensity)", type='p',pch='.')


g7stand<-vsn1[,17]
r7stand<-vsn1[,18]
m7<-r7stand-g7stand
a7<-(r7stand+g7stand)/2
plot(a7,m7,xlab="A7 (average of log pixel intensities)",ylab="M7 (difference in log pixel intensity)", type='p',pch='.')


g8stand<-vsn1[,19]
r8stand<-vsn1[,20]
m8<-r8stand-g8stand
a8<-(r8stand+g8stand)/2
plot(a8,m8,xlab="A8 (average of log pixel intensities)",ylab="M8 (difference in log pixel intensity)", type='p',pch='.')


g9stand<-vsn1[,21]
r9stand<-vsn1[,22]
m9<-r9stand-g9stand
a9<-(r9stand+g9stand)/2
plot(a9,m9,xlab="A9 (average of log pixel intensities)",ylab="M9 (difference in log pixel intensity)", type='p',pch='.')


g10stand<-vsn1[,23]
r10stand<-vsn1[,24]
m10<-r10stand-g10stand
a10<-(r10stand+g10stand)/2
plot(a10,m10,xlab="A10 (average of log pixel intensities)",ylab="M10 (difference in log pixel intensity)", type='p',pch='.')


g11stand<-vsn1[,25]
r11stand<-vsn1[,26]
m11<-r11stand-g11stand
```

```
a11<-(r11stand+g11stand)/2
plot(a11,m11,xlab="A11 (average of log pixel intensities)",ylab="M11 (difference in log
pixel intensity)", type='p',pch='.')


g12stand<-vsn1[,27]
r12stand<-vsn1[,28]
m12<-r12stand-g12stand
a12<-(r12stand+g12stand)/2
plot(a12,m12,xlab="A12 (average of log pixel intensities)",ylab="M12 (difference in log
pixel intensity)", type='p',pch='.')


g13stand<-vsn1[,29]
r13stand<-vsn1[,30]
m13<-r13stand-g13stand
a13<-(r13stand+g13stand)/2
plot(a13,m13,xlab="A13 (average of log pixel intensities)",ylab="M13 (difference in log
pixel intensity)", type='p',pch='.')


g14stand<-vsn1[,31]
r14stand<-vsn1[,32]
m14<-r14stand-g14stand
a14<-(r14stand+g14stand)/2
plot(a14,m14,xlab="A14 (average of log pixel intensities)",ylab="M14 (difference in log
pixel intensity)", type='p',pch='.')


g15stand<-vsn1[,33]
r15stand<-vsn1[,34]
m15<-r15stand-g15stand
a15<-(r15stand+g15stand)/2
plot(a15,m15,xlab="A15 (average of log pixel intensities)",ylab="M15 (difference in log
pixel intensity)", type='p',pch='.')


g16stand<-vsn1[,35]
r16stand<-vsn1[,36]
m16<-r16stand-g16stand
a16<-(r16stand+g16stand)/2
plot(a16,m16,xlab="A16 (average of log pixel intensities)",ylab="M16 (difference in log
pixel intensity)", type='p',pch='.')


g17stand<-vsn1[,37]
```

```
r17stand<-vsn1[,38]
m17<-r17stand-g17stand
a17<-(r17stand+g17stand)/2
plot(a17,m17,xlab="A17 (average of log pixel intensities)",ylab="M17 (difference in log
pixel intensity)", type='p',pch='.')


g18stand<-vsn1[,39]
r18stand<-vsn1[,40]
m18<-r18stand-g18stand
a18<-(r18stand+g18stand)/2
plot(a18,m18,xlab="A18 (average of log pixel intensities)",ylab="M18 (difference in log
pixel intensity)", type='p',pch='.')


g19stand<-vsn1[,41]
r19stand<-vsn1[,42]
m19<-r19stand-g19stand
a19<-(r19stand+g19stand)/2
plot(a19,m19,xlab="A19 (average of log pixel intensities)",ylab="M19 (difference in log
pixel intensity)", type='p',pch='.')


g20stand<-vsn1[,43]
r20stand<-vsn1[,44]
m20<-r20stand-g20stand
a20<-(r20stand+g20stand)/2
plot(a20,m20,xlab="A20 (average of log pixel intensities)",ylab="M20 (difference in log
pixel intensity)", type='p',pch='.')


#READING IN RAW DATA
tav1a<-read.table("C:\\Carletha\\TAV files\\new1atav.csv",sep=",",header=FALSE)
g1raw<-tav1a[,7]
r1raw<-tav1a[,8]

tav1b<-read.table("C:\\Carletha\\TAV files\\new1btav.csv",sep=",",header=FALSE)
g2raw<-tav1b[,7]
r2raw<-tav1b[,8]

tav1c<-read.table("C:\\Carletha\\TAV files\\new1ctav.csv",sep=",",header=FALSE)
g3raw<-tav1c[,7]
r3raw<-tav1c[,8]

tav1d<-read.table("C:\\Carletha\\TAV files\\new1dtav.csv",sep=",",header=FALSE)
g4raw<-tav1d[,7]
```

```
r4raw<-tav1d[,8]

tav2a<-read.table("C:\\Carletha\\TAV files\\new2atav.csv",sep=",",header=FALSE)
g5raw<-tav2a[,7]
r5raw<-tav2a[,8]

tav2b<-read.table("C:\\Carletha\\TAV files\\new2btav.csv",sep=",",header=FALSE)
g6raw<-tav2b[,7]
r6raw<-tav2b[,8]

tav2c<-read.table("C:\\Carletha\\TAV files\\new2ctav.csv",sep=",",header=FALSE)
g7raw<-tav2c[,7]
r7raw<-tav2c[,8]

tav2d<-read.table("C:\\Carletha\\TAV files\\new2dtav.csv",sep=",",header=FALSE)
g8raw<-tav2d[,7]
r8raw<-tav2d[,8]

tav4a<-read.table("C:\\Carletha\\TAV files\\new4atav.csv",sep=",",header=FALSE)
g9raw<-tav4a[,7]
r9raw<-tav4a[,8]

tav4b<-read.table("C:\\Carletha\\TAV files\\new4btav.csv",sep=",",header=FALSE)
g10raw<-tav4b[,7]
r10raw<-tav4b[,8]

tav4c<-read.table("C:\\Carletha\\TAV files\\new4ctav.csv",sep=",",header=FALSE)
g11raw<-tav4c[,7]
r11raw<-tav4c[,8]

tav4d<-read.table("C:\\Carletha\\TAV files\\new4dtav.csv",sep=",",header=FALSE)
g12raw<-tav4d[,7]
r12raw<-tav4d[,8]

tav8a<-read.table("C:\\Carletha\\TAV files\\new8atav.csv",sep=",",header=FALSE)
g13raw<-tav8a[,7]
r13raw<-tav8a[,8]

tav8b<-read.table("C:\\Carletha\\TAV files\\new8btav.csv",sep=",",header=FALSE)
g14raw<-tav8b[,7]
r14raw<-tav8b[,8]

tav8c<-read.table("C:\\Carletha\\TAV files\\new8ctav.csv",sep=",",header=FALSE)
g15raw<-tav8c[,7]
r15raw<-tav8c[,8]
```

```
tav8d<-read.table("C:\\Carletha\\TAV files\\new8dtav.csv",sep=",",header=FALSE)
g16raw<-tav8d[,7]
r16raw<-tav8d[,8]

tav12a<-read.table("C:\\Carletha\\TAV files\\new12atav.csv",sep=",",header=FALSE)
g17raw<-tav12a[,7]
r17raw<-tav12a[,8]

tav12b<-read.table("C:\\Carletha\\TAV files\\new12btav.csv",sep=",",header=FALSE)
g18raw<-tav12b[,7]
r18raw<-tav12b[,8]

tav12c<-read.table("C:\\Carletha\\TAV files\\new12ctav.csv",sep=",",header=FALSE)
g19raw<-tav12c[,7]
r19raw<-tav12c[,8]

tav12d<-read.table("C:\\Carletha\\TAV files\\new12dtav.csv",sep=",",header=FALSE)
g20raw<-tav12d[,7]
r20raw<-tav12d[,8]

#CREATES A MATRIX OF TWO ROWS AND ONE COLUMN OF PLOTS
#PLOTTING RAW DATA DUPLICATES AND STADARDIZED DATA
DUPLICATES ON SAME PAGE FOR COMPARISON

#GREEN DYE COMPARISONS
plot(g1raw,g2raw,xlab="Control 1 Raw Green (pixel intensity)",ylab="Control 2 Raw
Green (pixel intensity)", type='p',pch='.')
plot(g1stand,g2stand,xlab="Control 1 Stand. Green(pixel intensity)",ylab="Control 2
Stand. Green(pixel intensity)", type='p',pch='.')

plot(g3raw,g4raw,xlab="Treatment 1 Raw Green (pixel intensity)",ylab="Treatment 2
Raw Green (pixel intensity)",type='p',pch='.')
plot(g3stand,g4stand,xlab="Treatment 1 Stand. Green (pixel intensity)",ylab="Treatment
2 Stand. Green (pixel intensity), type='p',pch='.')

plot(g5raw,g6raw,xlab="Control 3 Raw Green (pixel intensity)",ylab="Control 4 Raw
Green (pixel intensity)", type='p',pch='.')
plot(g5stand,g6stand,xlab="Control 3 Stand. Green (pixel intensity)",ylab="Control 4
Stand. Green (pixel intensity),pch='.')

plot(g7raw,g8raw,xlab="Treatment 3 Raw Green (pixel intensity)",ylab="Treatment 4
Raw Green (pixel intensity)", type='p',pch='.')
plot(g7stand,g8stand,xlab="Treatment 3 Stand. Green (pixel intensity)",ylab="Treatment
4 Stand. Green (pixel intensity), type='p',pch='.')
```

```
plot(g9raw,g10raw,xlab="Control 5 Raw Green (pixel intensity)",ylab="Control 6 Raw Green (pixel intensity)", type='p',pch='.')
plot(g9stand,g10stand,xlab="Control 5 Stand. Green (pixel intensity)",ylab="Control 6 Stand. Green (pixel intensity), type='p',pch='.')

plot(g11raw,g12raw,xlab="Treatment 5 Raw Green (pixel intensity)",ylab="Treatment 6 Raw Green (pixel intensity)", type='p',pch='.')
plot(g11stand,g12stand,xlab="Treatment 5 Stand. Green (pixel intensity)",ylab="Treatment 6 Stand. Green (pixel intensity), type='p',pch='.')

plot(g13raw,g14raw,xlab="Control 7 Raw Green (pixel intensity)",ylab="Control 8 Raw Green (pixel intensity)", type='p',pch='.')
plot(g13stand,g14stand,xlab="Control 7 Stand. Green (pixel intensity)",ylab="Control 8 Stand. Green (pixel intensity), type='p',pch='.')

plot(g15raw,g16raw,xlab="Treatment 7 Raw Green (pixel intensity)",ylab="Treatment 8 Raw Green (pixel intensity)", type='p',pch='.')
plot(g15stand,g16stand,xlab="Treatment 7 Stand. Green (pixel intensity)",ylab="Treatment 8 Stand. Green (pixel intensity), type='p',pch='.')

plot(g17raw,g18raw,xlab="Control 9 Raw Green (pixel intensity)",ylab="Control 10 Raw Green (pixel intensity)", type='p',pch='.')
plot(g17stand,g18stand,xlab="Control 9 Stand. Green (pixel intensity)",ylab="Control 10 Stand. Green (pixel intensity), type='p',pch='.')

plot(g19raw,g20raw,xlab="Treatment 9 Raw Green (pixel intensity)",ylab="Treatment 10 Raw Green (pixel intensity)", type='p',pch='.')
plot(g19stand,g20stand,xlab="Treatment 9 Stand. Green (pixel intensity)",ylab="Treatment 10 Stand. Green (pixel intensity), type='p',pch='.')

#RED DYE COMPARISONS
plot(r1raw,r2raw,xlab="Treatment 1 Raw Red (pixel intensity)",ylab="Treatment 2 Raw Red (pixel intensity)", type='p',pch='.')
plot(r1stand,r2stand,xlab="Treatment 1 Stand. Red (pixel intensity)",ylab="Treatment 2 Stand. Red (pixel intensity)",type='p', pch='.')

plot(r3raw,r4raw,xlab="Control 1 Raw Red (pixel intensity)",ylab="Control 2 Raw Red (pixel intensity)", type='p',pch='.')
plot(r3stand,r4stand,xlab="Control 1 Stand. Red (pixel intensity)",ylab="Control 2 Stand. Red (pixel intensity)",type='p', pch='.')

plot(r5raw,r6raw,xlab="Treatment 3 Raw Red (pixel intensity)",ylab="Treatment 4 Raw Red (pixel intensity)", type='p',pch='.')
plot(r5stand,r6stand,xlab="Treatment 3 Stand. Red (pixel intensity)",ylab="Treatment 4 Stand. Red (pixel intensity)",type='p', pch='.')
```

```
plot(r7raw,r8raw,xlab="Control 3 Raw Red (pixel intensity)",ylab="Control 4 Raw Red (pixel intensity)", type='p',pch='.')
plot(r7stand,r8stand,xlab="Control 3 Stand. Red (pixel intensity)",ylab="Control 4 Stand. Red (pixel intensity)",type='p', pch='.')

plot(r9raw,r10raw,xlab="Treatment 5 Raw Red (pixel intensity)",ylab="Treatment 6 Raw Red (pixel intensity)", type='p',pch='.')
plot(r9stand,r10stand,xlab="Treatment 5 Stand. Red (pixel intensity)",ylab="Treatment 6 Stand. Red (pixel intensity)",type='p', pch='.')

plot(r11raw,r12raw,xlab="Control 5 Raw Red (pixel intensity)",ylab="Control 6 Raw Red (pixel intensity)", type='p',pch='.')
plot(r11stand,r12stand,xlab="Control 5 Stand. Red (pixel intensity)",ylab="Control 6 Stand. Red (pixel intensity)",type='p', pch='.')

plot(r13raw,r14raw,xlab="Treatment 7 Raw Red (pixel intensity)",ylab="Treatment 8 Raw Red (pixel intensity)", type='p',pch='.')
plot(r13stand,r14stand,xlab="Treatment 7 Stand. Red (pixel intensity)",ylab="Treatment 8 Stand. Red (pixel intensity)",type='p', pch='.')

plot(r15raw,r16raw,xlab="Control 7 Raw Red (pixel intensity)",ylab="Control 8 Raw Red (pixel intensity)", type='p',pch='.')
plot(r15stand,r16stand,xlab="Control 7 Stand. Red (pixel intensity)",ylab="Control 8 Stand. Red (pixel intensity)",type='p', pch='.')

plot(r17raw,r18raw,xlab="Treatment 9 Raw Red (pixel intensity)",ylab="Treatment 10 Raw Red (pixel intensity)", type='p',pch='.')
plot(r17stand,r18stand,xlab="Treatment 9 Stand. Red (pixel intensity)",ylab="Treatment 10 Stand. Red (pixel intensity)",type='p', pch='.')

plot(r19raw,r20raw,xlab="Control 9 Raw Red (pixel intensity)",ylab="Control 10 Raw Red (pixel intensity)", type='p',pch='.')
plot(r19stand,r20stand,xlab="Control 9 Stand. Red (pixel intensity)",ylab="Control 10 Stand. Red (pixel intensity)",type='p', pch='.')
```

APPENDIX D

R code for gene ongtology analysis

```
#all_goid freq analysis
data1<-read.table("C:\\Carletha\\GOID analysis
files\\all_goid.csv",sep=",",header=TRUE)
dim(data1)
#[1] 20076    89
data2<-data1[,1:87]
dim(data2)
#[1] 20076    87
trial<-rep(0,1746612)
data3<-as.matrix(data2)

for (i in 0:20075)
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}

data4<-sort(trial)
data5<-data4[1700000:length(data4)]

write.table(data5,"C:\\Carletha\\GOID analysis
files\\trial_allgoid.csv",row.names=TRUE,col.names=TRUE)



#t1_go prep in R
data1<-read.table("C:\\Carletha\\GOID analysis files\\t1_go.csv",sep=",",header=TRUE)
dim(data1)
#[1] 1010    88
data2<-data1[,2:88]
dim(data2)
#[1] 1010    87
trial<-rep(0,87870)
data3<-as.matrix(data2)
for (i in 0:1009)
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[50165:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\trial_t1go.csv",row.names=TRUE,col.names=TRUE)
```

```
#t2_go prep in R
data1<-read.table("C:\\Carletha\\GOID analysis files\\t2_go.csv",sep=",",header=TRUE)
dim(data1)
#[1] 1965    88
data2<-data1[,2:88]
dim(data2)
#[1] 1965    87
trial<-rep(0,170955)
data3<-as.matrix(data2)
for (i in 0:1964)
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[119148:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\trial_t2go.csv",row.names=TRUE,col.names=TRUE)
```

```
#t3_go prep in R
data1<-read.table("C:\\Carletha\\GOID analysis files\\t3_go.csv",sep=",",header=TRUE)
dim(data1)
#[1] 757    88
data2<-data1[,2:88]
dim(data2)
#[1] 757    87
trial<-rep(0,65859)
data3<-as.matrix(data2)
for (i in 0:756)
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[37514:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\trial_t3go.csv",row.names=TRUE,col.names=TRUE)
```

```
#t4_go prep in R
data1<-read.table("C:\\Carletha\\GOID analysis files\\t4_go.csv",sep=",",header=TRUE)
dim(data1)
#[1] 2129    88
data2<-data1[,2:88]
```

```
dim(data2)
#[1] 2129    87
trial<-rep(0,185223)
data3<-as.matrix(data2)
for (i in 0:2128)
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[129026:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\trial_t4go.csv",row.names=TRUE,col.names=TRUE)
```

```
#t5_go prep in R
data1<-read.table("C:\\Carletha\\GOID analysis files\\t5_go.csv",sep=",",header=TRUE)
dim(data1)
#[1] 2278    88
data2<-data1[,2:88]
dim(data2)
#[1] 2278    87
trial<-rep(0,198186)
data3<-as.matrix(data2)
for (i in 0:2277)
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[138018:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\trial_t5go.csv",row.names=TRUE,col.names=TRUE)
```

```
#t1_t5go prep in R
data1<-read.table("C:\\Carletha\\GOID analysis
files\\t1_t5go.csv",sep=",",header=TRUE)
dim(data1)
#[1] 328    88
data2<-data1[,2:88]
dim(data2)
#[1] 328    87
trial<-rep(0,28536)
data3<-as.matrix(data2)
for (i in 0:327)
```

```
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[19994:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\trial_t1_t5go.csv",row.names=TRUE,col.names=TRUE)



#control profile go prep in R
data1<-read.table("C:\\Carletha\\GOID analysis
files\\con_profile1_go.csv",sep=",",header=TRUE)
dim(data1)
#[1] 26    90
data2<-data1[,1:87]
dim(data2)
#[1] 26    87
trial<-rep(0,2262)
data3<-as.matrix(data2)
for (i in 0:25)
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[934:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\trial_conprofile1_go.csv",row.names=TRUE,col.names=TRUE)


data1<-read.table("C:\\Carletha\\GOID analysis
files\\con_profile2_go.csv",sep=",",header=TRUE)
dim(data1)
#[1] 53    90
data2<-data1[,1:87]
dim(data2)
#[1] 53    87
trial<-rep(0,4611)
data3<-as.matrix(data2)
for (i in 0:52)
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[1565:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\trial_conprofile2_go.csv",row.names=TRUE,col.names=TRUE)
```

```
data1<-read.table("C:\\Carletha\\GOID analysis
files\\con_profile3_go.csv",sep=",",header=TRUE)
dim(data1)
#[1] 57    90
data2<-data1[,1:87]
dim(data2)
#[1] 57    87
trial<-rep(0,4959)
data3<-as.matrix(data2)
for (i in 0:56)
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[889:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\trial_conprofile3_go.csv",row.names=TRUE,col.names=TRUE)


data1<-read.table("C:\\Carletha\\GOID analysis
files\\con_profile4_go.csv",sep=",",header=TRUE)
dim(data1)
#[1] 34    90
data2<-data1[,1:87]
dim(data2)
#[1] 34    87
trial<-rep(0,2958)
data3<-as.matrix(data2)
for (i in 0:33)
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[291:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\trial_conprofile4_go.csv",row.names=TRUE,col.names=TRUE)


data1<-read.table("C:\\Carletha\\GOID analysis
files\\con_profile5_go.csv",sep=",",header=TRUE)
dim(data1)
#[1] 6    90
data2<-data1[,1:87]
dim(data2)
#[1] 6    87
trial<-rep(0,522)
data3<-as.matrix(data2)
for (i in 0:5)
```

```
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[13:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\trial_conprofile5_go.csv",row.names=TRUE,col.names=TRUE)


data1<-read.table("C:\\Carletha\\GOID analysis
files\\con_profile6_go.csv",sep=",",header=TRUE)
dim(data1)
#[1] 46    90
data2<-data1[,1:87]
dim(data2)
#[1] 46    87
trial<-rep(0,4002)
data3<-as.matrix(data2)
for (i in 0:45)
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[1151:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\trial_conprofile6_go.csv",row.names=TRUE,col.names=TRUE)


data1<-read.table("C:\\Carletha\\GOID analysis
files\\con_profile7_go.csv",sep=",",header=TRUE)
dim(data1)
#[1] 41    90
data2<-data1[,1:87]
dim(data2)
#[1] 41    87
trial<-rep(0,3567)
data3<-as.matrix(data2)
for (i in 0:40)
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[1029:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\trial_conprofile7_go.csv",row.names=TRUE,col.names=TRUE)


data1<-read.table("C:\\Carletha\\GOID analysis
files\\con_profile8_go.csv",sep=",",header=TRUE)
```

```
dim(data1)
#[1] 10    90
data2<-data1[,1:87]
dim(data2)
#[1] 10    87
trial<-rep(0,870)
data3<-as.matrix(data2)
for (i in 0:9)
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[276:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\trial_conprofile8_go.csv",row.names=TRUE,col.names=TRUE)


#Master control profile list
data1<-read.table("C:\\Carletha\\GOID analysis
files\\final_con_go_profile.csv",sep=",",header=TRUE)
dim(data1)
#[1] 273    90
data2<-data1[,1:87]
dim(data2)
#[1] 273    87
trial<-rep(0,23751)
data3<-as.matrix(data2)
for (i in 0:272)
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[9972:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\allprofile_con_go.csv",row.names=TRUE,col.names=TRUE)


#trt profile go prep in R
data1<-read.table("C:\\Carletha\\GOID analysis
files\\trt_profile1_go.csv",sep=",",header=TRUE)
dim(data1)
#[1] 84    90
data2<-data1[,1:87]
dim(data2)
#[1] 84    87
trial<-rep(0,7308)
data3<-as.matrix(data2)
for (i in 0:83)
```

```
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[4360:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\trial_trtprofile1_goid.csv",row.names=TRUE,col.names=TRUE)


data1<-read.table("C:\\Carletha\\GOID analysis
files\\trt_profile2_goid.csv",sep=",",header=TRUE)
dim(data1)
#[1] 89    90
data2<-data1[,1:87]
dim(data2)
#[1] 89    87
trial<-rep(0,7743)
data3<-as.matrix(data2)
for (i in 0:88)
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[2611:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\trial_trtprofile2_go.csv",row.names=TRUE,col.names=TRUE)


data1<-read.table("C:\\Carletha\\GOID analysis
files\\trt_profile3_goid.csv",sep=",",header=TRUE)
dim(data1)
#[1] 31    90
data2<-data1[,1:87]
dim(data2)
#[1] 31    87
trial<-rep(0,2697)
data3<-as.matrix(data2)
for (i in 0:30)
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[316:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\trial_trtprofile3_go.csv",row.names=TRUE,col.names=TRUE)


data1<-read.table("C:\\Carletha\\GOID analysis
files\\trt_profile4_goid.csv",sep=",",header=TRUE)
```

```
dim(data1)
#[1] 65    90
data2<-data1[,1:87]
dim(data2)
#[1] 65    87
trial<-rep(0,5655)
data3<-as.matrix(data2)
for (i in 0:64)
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[1459:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\trial_trtprofile4_go.csv",row.names=TRUE,col.names=TRUE)



data1<-read.table("C:\\Carletha\\GOID analysis
files\\trt_profile5_goid.csv",sep=",",header=TRUE)
dim(data1)
#[1] 26    90
data2<-data1[,1:87]
dim(data2)
#[1] 26    87
trial<-rep(0,2262)
data3<-as.matrix(data2)
for (i in 0:25)
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[225:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\trial_trtprofile5_go.csv",row.names=TRUE,col.names=TRUE)



data1<-read.table("C:\\Carletha\\GOID analysis
files\\trt_profile6_goid.csv",sep=",",header=TRUE)
dim(data1)
#[1] 21    90
data2<-data1[,1:87]
dim(data2)
#[1] 21    87
trial<-rep(0,1827)
data3<-as.matrix(data2)
for (i in 0:20)
```

```
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[152:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\trial_trtprofile6_go.csv",row.names=TRUE,col.names=TRUE)



data1<-read.table("C:\\Carletha\\GOID analysis
files\\trt_profile7_goid.csv",sep=",",header=TRUE)
dim(data1)
#[1] 105    90
data2<-data1[,1:87]
dim(data2)
#[1] 105    87
trial<-rep(0,9135)
data3<-as.matrix(data2)
for (i in 0:104)
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[3398:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\trial_trtprofile7_go.csv",row.names=TRUE,col.names=TRUE)



data1<-read.table("C:\\Carletha\\GOID analysis
files\\trt_profile8_goid.csv",sep=",",header=TRUE)
dim(data1)
#[1] 26    90
data2<-data1[,1:87]
dim(data2)
#[1] 26    87
trial<-rep(0,2262)
data3<-as.matrix(data2)
for (i in 0:25)
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[107:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\trial_trtprofile8_go.csv",row.names=TRUE,col.names=TRUE)
```

```
#Master treatment profile list
data1<-read.table("C:\\Carletha\\GOID analysis
files\\final_trt_go_profile.csv",sep=",",header=TRUE)
dim(data1)
#[1] 447    90
data2<-data1[,1:87]
dim(data2)
#[1] 447    87
trial<-rep(0,38889)
data3<-as.matrix(data2)
for (i in 0:446)
 {for (j in 1:87)
 {trial[(87*i+j)]<-data3[(i+1),j]}}
data4<-sort(trial)
data5<-data4[24004:length(data4)]
write.table(data5,"C:\\Carletha\\GOID analysis
files\\allprofile_trt_go.csv",row.names=TRUE,col.names=TRUE)
final_trt_go_profile.csv
```