Report on the Molecular Investigations into the Jet Fuel and solvent exposure in the DeSeal/ReSeal programme conducted at the Mater Research Institute (UQ), Brisbane.

Professor FG Bowling

BSc, PhD, MBBS, FRCPA, FHGSA, MBA

31st JULY 2014

TABLE OF CONTENTS

PREFACE

(i) Confidentiality Statement

Property of Mater Research Institute (UQ)

This document is confidential and may not be used, divulged, published or otherwise disclosed without the consent of the Commonwealth of Australia, Department of Defence. The Intellectual property from research in this study is the property of Mater Research Institute (UQ).

(ii) How to read this report

For the lay/non-technical reader.

For the general background to the study and an overview of the study's structure, read Chapters 1 and 2. Chapters 3 and 4 can be skipped: they contain specifics about the recruitment of subjects, determination of exposure, measures, and analyses. For the results given in Chapters 3 the introduction and discussion sections can be read in isolation for a quick synopsis of the results in non-technical language. Detailed analyses are given in the Chapter 4 Appendices. An overall discussion of the entire results (i.e. how different findings "cluster" together) is presented in the Executive Summary Chapter 1.

(iii) Cover Image

Courtesy of Mater Health Services

Title: Research scientist in the laboratory, Mater Research Institute, Aubigny Place, Brisbane, Qld, 2014

(iv) Project Team Leaders

Professor Francis Bowling ***BSc, PhD, MBBS, FRCPA FHGSA, MBA*****.**

Professor Bowling is the Principal Investigator on the Jet Fuel Exposure Syndrome Study. He is currently the Director of Biochemical Diseases at Mater Children's Hospital and is a Professorial Fellow at the Mater Research Institute of the University of Queensland . Professor Bowling has an active research and clinical interest in the molecular basis of disease, particularly neurodegenerative diseases of children. He is a member of the Genomic Advisory Committee of the Royal Australasian College of Pathologists. He has developed a rigorous quality system for the validation of genomic technologies and monitoring of the diagnostic performance of clinical genomic laboratories. Professor Bowling is the Director of iMolecular Consulting, which offers expert opinion in molecular pathology.

Associate Professor Nigel Waterhouse ***BSc(Hons), PhD.***

After graduating from University College Dublin, Ireland in 1998, Associate Professor Waterhouse moved to the sunnier climes of Australia. With support from Dianne Watters and Martin Lavin at the Queensland Institute of Medical Research he completed his PhD on "Proteolysis in apoptosis", and then moved to Doug Green's laboratory at La Jolla Institute for Allergy & Immunology in San Diego to investigate how mitochondria regulate caspase activation in apoptosis. Nigel received a Peter Doherty Fellowship from the National Health and Medical Research Council (NHMRC) to return to Joe Trapani's Laboratory at the Peter MacCallum Cancer Centre in Melbourne to investigate the role of mitochondria in cell death induced by cytotoxic lymphocytes; cells of the immune system that protect the body from disease by seeking out and killing cancer or virus infected cells. This work continued with the support of an RD Wright Fellowship and a Career Development Award from the NHMRC, and afforded him the opportunity to begin his own group with the support of Ricky Johnstone and Grant McArthur in the Cancer Therapeutics Program at the Peter MacCallum Cancer Centre. The Apoptosis

and Natural Toxicity Laboratory was dedicated to understanding cancer cell death by cytotoxic drugs and natural killer cells, how cancer cells avoid cell death, and how we can use therapeutic drugs to kill cancer cells by restoring cell-specific defects in death pathways.

This theme is continued in Associate Professor Waterhouse's current research in the Apoptosis and Cytotoxicity Laboratory at the Mater Medical Research Institute which was started in 2009 with the support of an Australian Research Council Futures Fellowship and the Mater Foundation.

Dr Mervyn Thomas ***BA DPhil***

Mervyn Thomas is the founder and a partner in Emphron Research Pty Ltd: a biostatistics and bioinformatics consultancy. He has an international private practice, working with US and Australian biotech companies. His technical interests are in massively multivariate statistics for biomarker and diagnostic discovery and development. Before founding Emphron Research he led the biostatistics and bioinformatics research programme for CSIRO Mathematical and Information sciences. His career has involved senior roles in private and public sector research organisations. He is a member of the NHMRC research committee.

Professor Brett McDermott ***MD, FRANZCP***.

Brett McDermott is an Australian medical graduate and trained in Psychiatry and Child Psychiatry in the UK and Sydney. Current appointments are the Executive Director of the Mater Child and Youth Mental Health Service in Brisbane, Chair of the Queensland Child and Youth disaster response, Professorial Fellow at the Mater Medical Research Institute and Adjunct Professor at the Queensland University of Technology. Professor McDermott is also a By-Fellow at Churchill College, Cambridge University; and a Director of the Australian national depression initiative, *beyondblue*. Recently, (2013) Professor McDermott was appointed Director of the new adolescent and young adult hospital at the Mater (MAYAC). Clinical and research areas of

interest include children and adolescents with depression, post-traumatic mental health, and child and youth mental health service provision.

Professor Deon Venter ***MBChB, PhD, MBA, FRCPA***

Professor Venter is Director of Pathology, Mater Health Services; Professor of Pathology, School of Medicine, University of Queensland; Honorary Principal Research Fellow Mental Health Research Institute, Melbourne; Honorary Consultant Pathologist, Victorian Clinical Genetics Services Ltd, Melbourne; and Founder and Medical Advisor, Athlomics Pty Ltd.

He is a consultant anatomical and molecular pathologist who has driven the establishment of a genomic and proteomic diagnostics facility at Mater Pathology. Professor Venter has been directly responsible for the launch of several new gene-based diagnostics in Australian companies and reference laboratories. He has acted as a consultant to major pharmaceutical companies in addition to biotechnology companies based in Europe and the USA.

Professor Venter has served on major global bodies such as the World Health Organisation, where he helped to formulate strategies to implement genetic services in developing countries and the Expert Group of the International Olympic Committee (IOC) Medical Commission on the molecular basis of connective tissue and muscle injuries in sport. Professor Venter is an author of over 90 publications on cancer genetics and genetic-based diagnostics.

Project Team Members

The contribution of each of the following members of the research team is acknowledged:-

Avis McWhinney, ***B Med Lab Sci*** is the Supervising Scientist of the Metabolic Section, Department of Clinical Chemistry, Mater Pathology, Brisbane Australia and has over 25 years experience in this area and who developed and validated the HPLC, UPLC, LC-MS/MS and GCMS assays .

Elisa Jansen, ***PhD***, Cell Biologist, Mater Research Institute.

Lisa Crowley, ***PhD***, Cell Biologist, Mater Research Institute.

Melinda Christensen, ***PhD***, Cell Biologist, Mater Research Institute.

Adrian Scott, ***PhD***, is a molecular biologist at the Mater Research Institute who worked on development of the cell exposure model.

Gareth Price, ,***PhD*** supervised the genomic studies at Mater Pathology.

Renee Gallagher, ***B. App Sci***, is a cytogeneticist in Mater Pathology who performed the genetic studies using Array technology.

Tristan Wallis ***PhD*** is a biochemist at Mater Pathology who developed the plasma and cell proteomic analysis.

Megan Bowlay, ***PhD*** is a biochemist at Mater Pathology who performed plasma and cellular proteomic analysis by MALDI-TOF.

Brooke Marfell, ***PhD***, is a cell biologist at Mater Research Institute, who undertook cell exposure studies of the toxicity of compounds and the mechanism of that toxicity.

Rachelle Warner, is a toxicologist in the Defence Department.

ACKNOWLEDGEMENTS

The project team also wish to acknowledge the support of :-

Professor Ian R. Gardner ***MBBS, MPH, FAFOEM, FACOEM, FRSM***

Senior Physician in Occupational and Environmental Medicine, Defence Centre for Occupational Health.

Professor Mark Bowles ***PhD GradDipMgt***

Chief Operating Officer, Mater Research Institute.

Ms Nicole Shiveley

Deputy Director, Research Management Office UQ

Representatives of the JFES Study Consultative Forum Committee:

Hon Arch Bevis - Chairman

Professor Ian R. Gardner - Senior Physician in Occupational and Environmental Medicine, Defence Centre for Occupational Health

Mr Max Chambers – DeSeal/ReSeal Support Group

Mr Geoff Curl – DeSeal/ReSeal Support Group

Mr David Janik – Department of Veterans Affairs

Wing Commander Ron Usher (Retd) – RAAF Association

Mr Viv Quinn – RSL Australia (Qld)

WOFF-AF Mark Pentreath – CDF Representative

Wing Commander Lance Halvorson (Retd) – RAAF Association

Commodore Michael Dowsett (Retd) – Defence Force Welfare Association

EXECUTIVE SUMMARY

Overview

The main objective of this project was to investigate the toxicity of JP-8 fuel and the solvents used in the Deseal/Reseal programme using a systems biology approach. In the exposure environment (fuel tanks, aircraft hangers etc), workers were typically exposed either by inhalation of vapours or by absorption through the skin. There were occasionally reports of direct ingestion through the mouth. Health studies of exposed workers and other research reports show premature death for some individuals, an increased risk of unusual malignancies in internal organs such as small bowel, erectile dysfunction, and behavioural disturbances. These findings may manifest years after exposure suggesting changes to the cells and tissues not directly exposed to the fuel and solvents. Changes to the systems biology was investigated by proteomic and genomic studies.

Laboratory cell studies of DeSeal/ReSeal compounds

Development of Cell exposure model

Previous methods for studying cellular responses to JP8 and solvents involved direct addition of these compounds to cells in laboratory growth plates using other solvents such as Ethanol. These methods were considered to be inadequate because they did not recognise the role of circulating blood plasma in distributing these compounds to internal organs. The JFES project team developed a method of studying cells by exposing them to blood plasma, which they believe is a better model of the inhalation and skin exposure routes for distributing solvents to internal organs. This method has been published in a peer reviewed journal. (See Appendix 1)

Distribution of JP8 and DeSeal/ReSeal solvents

The studies of plasma exposed to JP8 and solvents showed that the compounds are not distributed by plasma in the same proportions as found in the fuel and solvent mixtures. This means that higher levels of some compounds are actually presented to cells and organs than those

proportions in the fuel solvent mixtures. The study showed that the majority of the compounds are distributed by binding to plasma lipids rather than simply dissolved in the plasma water. This raises the possibility that individuals with higher bloods lipids may distribute more of the compounds to internal organs.

The effects of the JP8 and solvents on cells

The study then tested the effects of the JP8 and solvents on cells. The JP8 and solvents were tested as a mixture and individually. The key findings were:-

Plasma exposed to JP8 alone is directly toxic to cells

Plasma exposed to a mixture of JP8 combined with solvents has greater toxicity to cells with 40% cells showing changes before 4 hours, and 90% cells affected at 12 hours.

The following individual components were found to have the highest cellular toxicity:-

- Kerosene
- Benzene and butylbenzene
- All Alkanes including iso-octane, decane, dodecane, tetradecane and hexadecane
- Diegme
- N, N Dimethyl acetimide
- Naptha
- Thiophenol

The solvents used in the Deseal/Reseal programme demonstrated either low cell toxicity or manifest toxicity to a lesser extent than the JP8 fuel components.

Effects on gene expression

Gene expression in cells was altered following exposure. Changes greater then 5 fold were considered significant. The genes altered are shown in Table (3). The function of these genes involved mostly cell survival/death, metabolism, cell cycle, DNA maintenance (housekeeping), and cell regulation. These genes have been implicated in pathological processes including cancer, neurodegeneration, and immune suppression.

Effects on proteins

Cellular proteins were altered after exposure. The changes to cellular proteins reflected the changes in gene expression involving cell survival/death, metabolism, cell division, and roles in cellular gene transcription/translation.

Cell Death

Cell death occurred by two mechanisms. A number of cells appeared more vulnerable with death occurring by disruption of cellular membranes and by lysis (bursting) of the cells. The more common mechanism of cell death was by apoptosis, which is a programmed response of cells to injury. Not all injured cells undergo complete apoptosis indicating persistence of injured cells. This may suggest a survival of injured cells with malignant potential. The cell culture methods could not determine the long term effects.

Study of exposed workers

The study of exposed workers showed differences from the matched control group in health indices, and in some genomic studies. The changes were not as significant as those seen in the acute cell exposure model in the laboratory.

Rating of exposure

Because of the unavailability of accurate exposure data (degree and duration), a problem also encountered in other studies, the workers were classified into 3 groups.

1. Definite high exposure who worked inside the fuel tanks
2. Significant contact such as by dosing of skin or accidental ingestion
3. Minimal contact in the general area such as collection of rags or cleaning of the area.

Health Assessment Scores

The Health assessment scores showed exposed workers to have a lower health rating than controls. There did not appear to be a decrease in the health scores (dose response) related to the degree of exposure. Workers with mild exposure had the same decrease in their health scores as those with high exposure. This suggests that other factors beyond the Deseal/ Reseal contact have decreased the health scores.

Genetic studies of blood cells from exposed workers

All studies were undertaken on plasma and white blood cells as these were the only tissues for which it was possible to obtain samples. The genetic studies of blood cells examined two types of changes in gene expression, the presence of chromosomal changes, and for appearance of mutations in the mitochondrial DNA. There were no chromosomal changes detected at a level of 50Kb using a high resolution SNP ARRAY.

There were no changes in the mitochondrial DNA mutation load between exposed workers and age matched controls (Mitochondrial DNA changes can accumulate with age).

There were no changes in the amount or type of protein coding mRNA expression, which is an index of cell activity. In disease states , these are usually tissue specific and may not appear in blood cells unless they are directly involved in the disease process.

There were small but significant and consistent changes in the expression of regulatory microRNAs that control activity of other genes. The regulatory functions of the altered genes have been linked to neurological changes and neurodegenerative disorders. It must be emphasised that interpretation of the function of regulatory genes is an evolving science with much uncertainty at present. The regulatory genes, which compose 98% of our genome, have a major role in human development, adaptation and response to disease. The function is only known for ~40% of these at present. Disease causing associations, with some early exceptions, are still unmapped.

Protein studies of plasma and blood cells

No significant changes were seen in the levels and types of protein expressed in the plasma and blood cells of exposed workers. A few small changes were seen consistently, but these did not reach a level that the researchers considered significant.

Discussion and Conclusions

Confounders and sensitivity

Dose response not detected

A dose response would have been expected but was not observed in the workers with different exposure histories. The unexpected similarity in the health scores and genomic studies within the exposed groups (low, medium, high) raises several hypotheses:-

Confounders

There are other factors independent of Deseal/Reseal exposure which could produce the changes seen. Confounders could include:-

- An ascertainment bias whereby only those workers *affected* by *any* exposure volunteered to participate in the study.
- An ascertainment bias whereby only those workers NOT affected by the exposure (i.e.. Survivors) volunteered to participate in the study.
- The workers were stratified by their exposure to Deseal/Reseal materials. The effects seen may NOT be due to the Deseal/Reseal materials but to some other experience of the workers. The cellular studies suggest that exposure to fuel alone could be responsible.
- It was not possible to examine other possible shared confounding events in the work careers or in the lifestyle of the personnel. (e.g.. other occupational exposure not related to Deseal/Reseal such as medications, substance abuse, nutrition)
- This study was conducted on individuals between 10 and 30 years after their exposure. If significant changes occurred at the time of exposure, normal cellular repair and selection mechanisms may have lessened the biological signal that could be observed in this study. The small but consistent changes observed suggest this possibility. Either the effect at the time was minimal but has persisted, or the effect was larger but has diminished over-time.

- The cellular studies show that the compounds are mostly distributed by plasma lipids. The exposure to organs within the body would likely depend on the concentration of plasma lipids at the time of fuel exposure. Plasma lipids vary genetically between individuals, with lifestyle and alcohol intake, with composition of their diet, as well as the time after meals when the exposure occurred. The lack of a dose effect could be explained if workers in the lower exposure group had higher plasma lipids at the time of exposure. Individuals in the high exposure group worked within the fuel tanks and were selected because they were leaner and smaller, possibly protected to some extent by lower plasma lipids.

Significance of findings

The cellular findings, supported by other recently published genomic studies, indicate a definite toxicity from JP-8 to exposed cells . The components of JP8 tested are commonly found in most (aviation) fuels. The results indicate that there is a need for concern about exposure to fuels in general.

The study was not designed to determine the degree of occupational exposure necessary to produce cellular changes. However, the results show that cells grown in a nutrient containing as little as 5% exposed plasma are affected. In the body, blood cells have 100% exposure to plasma while other organs will have less exposure depending on the net blood flow and cellular membrane barriers. Organs such as brain, liver and bowel have very high blood flow. Cellular membranes generally have greater permeability to substances dissolved in lipids.

The study was also not designed to determine the most toxic routes of exposure (inhalation, ingestion, skin contact), but did demonstrate that fuel components can be distributed to organs through blood plasma, i.e. organs such as brain or liver, not directly exposed in the contact, may undergo secondary exposure. The implication is that all body systems must be considered in assessing/monitoring the health of exposed workers.

While the changes seen many years after exposure were small, they were consistent. The changes are most apparent in gene regulation and had some association to the health problems (e.g., malignancy) identified in other studies.

There were no chromosomal changes or mutations linked to the exposure. The genes changes seen can be described as Epigenetic, which is a mechanism of cellular adaptation to some environmental influence. Epigenetic changes are less clearly linked (at the present knowledge) to disease. Epigenetic changes occur through a variety of cellular mechanisms and these were not investigated in this study. Some epigenetic changes can be transferred down through successive generations but currently have not been shown to cause birth defects or mutation in off-spring.

Recommendations

The cell results show a definite cellular toxicity from JP8 fuel. The components of the fuel exhibiting toxicity are common to most fuels. Consideration should be given to further studies of workers exposed to fuel of any type.

Newer genomic and bioinformatic technologies have been developed during the time of this study and have been employed in other studies of occupational fuel exposure. These technologies can be applied to other exposure risks (including PTSD) in defence (veteran) health risk assessment. An expert committee should be constituted to advise on research and clinical application of these technologies.

Plasma free DNA sequencing can now be used to assess (from blood samples), the cellular death associated with tumours , transplant rejection, miscarriage, and infections. Targeted RNA expression studies can reveal immediate changes in gene activity following fuel exposure. A study of workers with recent or past fuel exposure is recommended.

The best time to study cellular changes would be immediately after direct exposure. A protocol should be established for assessment of an exposed individual to include sample collection immediately after the exposure for quantification of plasma lipids, plasma fuel components, free DNA sequencing, and targeted RNA expression.

Exposed veterans should be reassured that while small and consistent changes were observed in this study, there were no changes detected known to have immediate or severe health consequences. The changes support the findings from other studies that there is a possible increased risk of developing health problems. As the changes observed are in gene regulation, it is also possible that healthy lifestyle changes may ameliorate the risk.

CHAPTER 1

Background and Overview of Research Plan

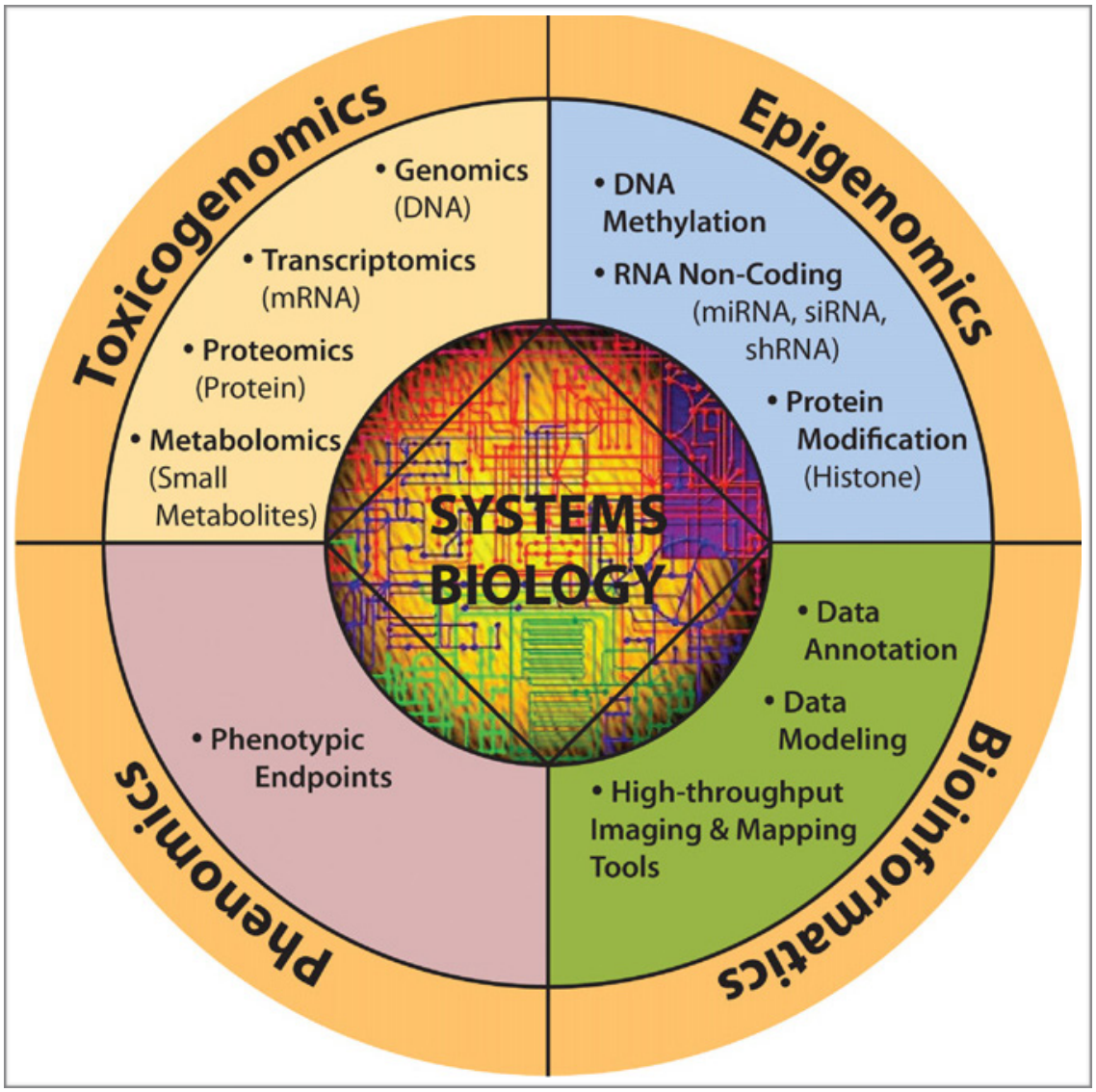


Fig. 1. Overview of systems biology and its components (centre of image, Wired Systems Biology, adapted from Chemical & Engineering News 81 [20) 2003). (methylomics), non-coding microRNA (miRNAomics) along with small interfering RNA (siRNA) and short hairpin RNA (shRNA) for RNA interference (RNAi), and histone modification. Epigenetic modifications play an essential role in regulating gene expression and biological and molecular functions in living cells, without altering the genome. Another tool central to systems biology is bioinformatics, the application of computational information technology to the field of molecular biology to understand how cells and cell systems work [6,7]. Bioinformatics facilitates the analysis of complex biological data (toxicogenomic and epigenomic endpoints) and applies knowledge from annotated functions, pathways and networks to describe normal and perturbed biological states, also known as phenomics, the study of outcomes (phenotypic endpoints). Together, these omic technologies can each provide a "molecular signature" or "fingerprint" of chemical exposure, early effect or genetic susceptibility, which may enhance our understanding of gene–environment interactions. Thus, this approach has the potential to comprehensively define the mechanisms contributing to disease.

1. INTRODUCTION

1.1 Overview

Systems biology is a recent and evolving interdisciplinary field that focuses on the systematic study of complex interactions in biological systems. Systems biology employs a multidisciplinary approach to study all components and interactions in the network of DNA (genes), RNA, proteins and biochemical reactions within a cell or organism. This new field utilises powerful tools that include toxicogenomics, epigenomics, bioinformatics, and phenomics, classical toxicological or phenotypic endpoints (Fig. 1). Toxicogenomics combines toxicology with molecular profiling technologies, including genomics (DNA), transcriptomics (mRNA), proteomics (proteins) and metabolomics (chemical metabolites) to elucidate molecular mechanisms involved in chemically induced toxicity. Chemically induced alterations in the transcriptome, proteome, and metabolome are analysed in the context of the stable, inherited genome, which is assessed by genomics. Toxicogenomic studies of human populations are crucial to understanding gene–environment interactions, and can provide the ability to develop novel biomarkers of exposure, early effect, and susceptibility (genome). Epigenomics is the study of epigenetic elements, including DNA methylation, but also histone modification. Alkylation of histone proteins has been shown following hydrocarbon exposure.

1.2 Application of systems biology in studies of benzene and hydrocarbon toxicity

The systems biology approach has been informative in another incident of benzene exposure involving chemical plant workers. Toxicogenomic studies, including genome-wide analyses of susceptibility genes (genomics), gene expression (transcriptomics), protein expression (proteomics), and epigenetic modifications (epigenomics), of human populations exposed to benzene were used to understanding gene–environment interactions, providing the ability to develop biomarkers of exposure, early effect and susceptibility.

Comprehensive analysis of toxicogenomic and epigenomic profiles by bioinformatics in the context of phenotypic endpoints, helped define the mechanisms by which benzene causes leukaemia. A systems biology approach was applied to a molecular epidemiology study of workers exposed to benzene. Hematotoxicity, defined as a significant decrease in almost all blood cell counts, was identified as a phenotypic effect of benzene that occurred even below 1ppm benzene exposure. A significant decrease in the formation of progenitor colonies arising from bone marrow stem cells with increasing benzene exposure, showed that progenitor cells are more sensitive to the effects of benzene than mature blood cells, likely leading to the observed hematotoxicity. Analysis of transcriptomics by microarray in the peripheral blood mononuclear cells of exposed workers, identified genes and pathways (apoptosis, immune response, and inflammatory response) altered at high (>10 ppm) and low (<1 ppm) benzene levels. Serum proteomics by SELDI-TOF-MS revealed proteins consistently down-regulated in exposed workers. Preliminary epigenomics data showed effects of benzene on the DNA methylation of specific genes.

Benzene, a ubiquitous chemical, is an established cause of acute myeloid leukaemia (AML), myelodysplastic syndromes (MDS), and probably lymphocytic leukemias and non-Hodgkin lymphoma (NHL) in humans. Benzene-induced toxicity in blood forming systems has been known for more than a century. In 1982, the International Agency for Research on Cancer (IARC) stated, “There is sufficient evidence that benzene is carcinogenic to man”, and when a new IARC classification system was established in 1987, benzene was immediately placed in the Group 1 human carcinogen category·· Potential mechanisms of benzene toxicity have been investigated primarily in the following areas: (1) benzene metabolism in the liver (CYP2E1, etc.) and transportation to the bone marrow for secondary metabolism (MPO, NQO1) ; (2) oxidative stress from reactive oxygen species generated by redox cycling ; (3) chromosome alterations including translocations, deletions, and aneuploidy; (4) protein damage to tubulin, histone proteins, topoisomerase II, etc. and (5) immune system dysfunction (TNF-_, INF-_, AhR, etc.).

Benzene induces chromosomal alterations similar to those found in therapy-related MDS and AML (t-MDS/AML), and in de novo leukaemia. Distinct chromosome effects arise following exposure to alkylating agents (5q-/-5 or 7q-/-7 and associated genetic abnormalities) and topoisomerase II inhibitors (recurrent balanced translocations or inversions) used in chemotherapy treatment. Exposure to benzene or its metabolites has been associated with loss and long (q) arm deletion of chromosomes 5 and 7 and translocations involving t(21q), further suggesting that benzene induces leukaemia through multiple different mechanisms. Studies to date have provided evidence for multiple potential mechanisms using approaches that rely on limited research tools that analyse only one or a few, a priori selected genes, pathways or metabolites at a time. A systems biology approach is necessary to interrogate all potential mechanisms by which benzene and other hydrocarbon exposures contribute to disease.

1.3 Gene expression profiling by transcriptomics

Transcriptomic studies are useful in determining the impact of environmental or occupational exposure to chemicals on the transcriptome, the set of all mRNA transcripts expressed within a cell. To better understand the risks of benzene in humans, the peripheral blood mononuclear cell (PBMC) transcriptomes from occupationally exposed workers in China were examined by microarray (Affymetrix). Analysis of six exposed-control pairs revealed differential expression in 29 genes in the exposed individuals, compared to the controls. Four genes, CXCL16, ZNF331, JUN, and PF4, were shown to be potential biomarkers of early response to benzene exposure as they were confirmed by quantitative-polymerase chain reaction (q-PCR). A later study of eight exposed-control pairs confirmed these results, using two different microarray platforms (Affymetrix and Illumina) to identify global gene expression changes. The differential expression of 2692 genes and 1828 genes was found by Affymetrix and Illumina, respectively, and the 4 genes, CXCL16, ZNF331, JUN, and PF4, were among the most significantly altered, validating the findings from the earlier Forrest et al. study.This study

additionally identified biological pathways that were associated with high benzene exposure, including genes involved in apoptosis and lipid metabolism. The effects of exposure to high levels of benzene are well documented compared to low-level exposure, the latter being more challenging due to confounders. More recently, in an expanded study of 125 factory workers, low-dose benzene exposure (<1 ppm, n = 59) can also cause widespread subtle, yet highly significant, perturbation of gene expression in PBMC. The study was designed with sufficient power to detect robust expression changes, accounting for technical variability as well as age, gender and other confounders. Microarray analysis revealed significant dysregulation of more than 2500 genes by low-dose benzene exposure, over 70 of which had differential expression ratios exceeding 1.5.

Several of the detected genes exhibited significantly altered expression only at low levels of benzene exposure, and are thus potential biomarkers of low-dose exposure. The findings show that even low levels of occupational benzene exposure cause a significant perturbation of expression of genes involved in immune and inflammatory responses. These techniques will be employed in this study.

1.4 Proteomic biomarkers of benzene and hydrocarbons

Another important toxicogenomic tool, proteomics, can be used to measure alterations in the proteome (e.g., protein levels, post-translational modifications) associated with exogenous chemical exposure. Effects on the blood proteome may reflect effects at distal body sites. As with transcriptomics, proteomics can be used to discover biomarkers of exposure and early effect, as well as increase our understanding of the mechanisms underlying disease. The impact of benzene on the human serum proteome in exposed factory workers and controls was used to obtain insight into the mechanism of action of benzene. Serum samples were fractionated and proteins were bound to surface-enhanced laser desorption/ionisation time-of-flight mass spectrometry (SELDI-TOF- MS) chips. Three proteins (4.1, 7.7, and 9.3 kDa) were consistently down-regulated in the exposed (n = 10) compared to the control (n = 10) individuals in two separate sets of study subjects (40 subjects total). All proteins were highly inversely correlated with individual estimates of benzene exposure. The 7.7- and 9.3-kDa proteins were identified as platelet factor 4 (PF4), also down-regulated at the gene expression level described above, and connective tissue activating peptide CTAP-III, respectively, both platelet-derived CXC chemokines. Thus, reduced protein levels of PF4 or CTAP-III are potential biomarkers of the early biologic effects of benzene.

Our proteomic studies could identify further biomarkers of JP8 exposure, and elucidate the mechanisms underlying toxicity and associated disease. A similar mass spectrometry approach using MALDI-TOF and LC-Electrospray will be employed in this study.

1.5 Epigenomics in pilot benzene studies

Gene expression and ultimately protein expression is regulated at the epigenetic level by processes including DNA methylation, histone modification and miRNA (microRNA) expression. The epigenome, while stable through cell division and even in some cases reproduction, can be reprogrammed by nutritional, chemical, and physical factors. Further, epigenetic modifications represent more stable biomarkers and fingerprints of exposure than altered gene or protein expression. Thus, the study of effects on the epigenome is crucial to understanding mechanisms of toxin action. While epigenetics refers to the study of individual or specific gene activity, epigenomics focuses on global analyses of epigenetic changes across the entire genome.

A recent study reported that hyper-methylation in p15 and genome-wide hypo-methylation assessed by LINE-1 (Long Interspersed Nuclear Element-1) were associated with very low benzene exposures (<22 ppb), in healthy subjects including petrol station attendants and traffic police officers, although the corresponding effects on methylation were very low. To determine whether epigenetics plays a role in the hematotoxicity of benzene, recent pilot epigenomic studies including DNA methylation and miRNA expression arrays identified changes in the blood of workers occupationally exposed to benzene.

Hematotoxicity is a phenotypic outcome of benzene exposure. Although benzene was known to have toxic effects on the hematopoietic system (hematotoxicity) at high, occupational doses for over a century, the degree of hematotoxicity at low levels of exposure was largely unknown. Recently, a study of 250 workers exposed to varying levels of benzene and 140 unexposed controls in Tianjin, China, during which benzene and other chemical exposure levels were monitored repeatedly for up to 12 months, was conducted. Air, urine and blood samples were collected and complete blood counts (CBC) analysed. In comparison with the non-exposed controls (n = 140), a significant decrease was observed in almost all blood cell counts, such as white blood cells (WBC), granulocytes, lymphocytes and platelets, in

workers exposed to benzene (n = 250), even at exposures below 1ppm (n = 109), the current occupational standard in the U.S. Additionally, lymphocyte subset analysis showed significant, dose-dependent, decreases in CD4+-T cells, CD4+/CD8+ ratio, and B cells at <1ppm benzene exposures. These findings, based on the differentiated blood cell counts, provide evidence of bone marrow toxicity in workers exposed to benzene at or below 1ppm. Because all types of WBC counts were suppressed, it was suspected that the number or functionality of hematopoietic stem and/or progenitor cells generated in the bone marrow had been reduced by benzene.

To test this hypothesis, progenitor cells circulating in peripheral blood were examined to see the effects of benzene on different types of progenitor cell colony formation (CFU-GM, BFU-E, CFU-GEMM). The results showed highly significant dose-dependent decreases in colony formation from all three types of progenitor cells, especially when compared to the corresponding decreased levels of differentiated WBC and granulocytes[27]. This suggests that early myeloid progenitor cells are more sensitive than mature cells to the effects of benzene, and clarifies the role of benzene in reduced blood cell counts. Overall, these haematological effects could reflect events in the bone marrow that may be associated with adverse health effects in the future.

Having established that hematotoxicity, specifically effects myeloid progenitor cells, as a phenotypic anchor of benzene toxicity, the molecular mechanisms underlying these effects, through the comprehensive systems biology approach should be undertaken to include the JP8 fuel and solvents in Deseal/Reseal exposure. The MRI laboratory has special expertise in Dendritic Cell analysis, prime mediators of haematological response.

1.6 Historical Background

Purchased from the United States of America in the 1970s, the F-111 strategic strike aircraft does not have a dedicated fuel bladder, with fuel instead occupying the empty spaces between other metal structures. Over time, the fuel-tight sealant between mating and internal structures degraded and it was necessary to carry out periodic seal repair programs. In most cases, this involved firstly removing the original sealant (desealing) and then replacing it with new sealant (resealing), with one program spraying new sealant over the existing defective sealant. In Australia, four formal F-111 fuel tank Deseal/Reseal (DSRS) programs were implemented over more than two decades (1975-1999), each involving different processes and the use of approximately 60 hazardous substances, mainly organic solvents. Workers on the DSRS programs frequently spent extended periods of time inside the fuel tanks in conditions that were cramped and inadequately ventilated. Throughout this period concerns had been raised about the health of DSRS workers, culminating in a Board of Inquiry appointed by the Chief of Air Force in July 2000. Concurrently it was decided to conduct an epidemiological study of the health outcomes of F-111 DSRS workers. Between 2001 and 2004, a comprehensive, large-scale epidemiological investigation was conducted on behalf of the Australian Department of Defence into the health of all personnel involved in DSRS activities in Australia since the programs began.

Three reports from the epidemiological studies were published:

The neurological study showed there was a strong and consistent increase in self-reported cognitive problems among the exposed, and small but significant exposure-specific differences on objective tests of cognitive functioning. The findings are consistent with the development of adverse neuropsychiatric changes related to occupational exposure to solvents.

A further study of mental health showed there was high agreement between self-reported mood disturbance and objective tests. The exposed group was more likely to self-report previous diagnoses of depression/anxiety, had higher use of antidepressant medications, and had increased risk of

diagnosis of depression/anxiety. Results were consistently strong against both comparison groups, with the exposed more likely to have mental distress and social dysfunction when compared with the Australian population. The study showed there is robust evidence for an association between F-111 DSRS exposure and impaired mental health.

A study of cancer incidence and mortality in aircraft maintenance workers showed cancer incidence was higher in the exposed group, with marginally significant increases of 40-50% (cancer incidence rate ratio range 1.45-1.62). Exposed group mortality was significantly lower than both comparison groups, likely due to survivor bias in the exposed group (mortality rate ratio range 0.33-0.44). The study showed that, on the balance of probabilities, there is an increased risk of cancer associated with participation in F-111 DSRS activities.

Chapter 2

Study of JP8 fuel and Deseal/Reseal solvents in a cell toxicity model

Development of exposure model for cell studies

Exposure to jet propulsion-8 (JP-8) aviation fuel and chemicals used in aircraft maintenance have been linked to a variety of pathologies that may occur as a consequence of their systemic effect on cells in the body. Using in vitro culture models to understand these effects poses a significant problem because JP-8 is largely insoluble in aqueous media that is used to culture cells. Previous studies have attempted to overcome this by dissolving the JP-8 in ethanol but this approach is not physiologically robust because ethanol is likely to deliver components of JP-8 that may not contribute to adverse health effects in vivo. Components of JP-8 that act systemically are likely to be distributed to organs by plasma regardless of the route of exposure. The components and relative concentrations of the components of JP8 are listed in the Table 1. We therefore investigated whether human plasma could be used as a physiologically relevant carrier for JP-8 to test its effects on cells in vitro. JP-8 did not dissolve in human plasma but human plasma that had been mixed with JP-8 was toxic to cultured cells in a dose dependent manner. This effect was specific because serum free media, water or serum albumin mixed with JP-8 was not toxic when added to cells. The method we have developed and published which exposes human plasma to JP-8 and solvents represents a new more physiologically relevant strategy for delivery of JP-8 to cells in vitro that will facilitate research into understanding the effects of occupational exposure.

The physical properties of JP-8 such as water insolubility and volatility are a significant challenge to in vitro toxicological studies to investigate the effects of occupational or environmental exposure.

Table 1. Molecular compounds and relative concentrations forming the JP8 fuel mix

	Component	Proportion (%)	Solubility	Molar Mass	Formula	Use	Molarity
	DiEGME	0.15	Miscible with water	120.2 g/mol	$C_5H_{12}O_3$	De-icing additive for jet fuel.	12.5 mM
	Kerosene	99.9	Insoluble in water; Miscible with other petroleum solvents				
Kerosene	Benzene	>0.02	Miscible with alcohol, chloroform, ether, carbon disulfide, carbon tetrachloride, glacial acetic acid, acetone, oils	78.11 g/mol^{-1}	C_6H_6	Gasoline additive	> 2.6 mM
	Methylcyclohexane	3.51	Insoluble in water at 20°C	98.186 g/mol	C_7H_{14}	Solvent	358 mM
	Isooctane	3.66	Practically insol in water; somewhat sol in abs alcohols; sol in benzene, toluene, xylene, chloroform, ether, carbon disulfide, carbon tetrachloride, DMF and oils, except castor oil.	114.23 g/mol^{-1}	C_8H_{18}	Solvent and thinner	320 mM
	m-Xylene	3.95	Miscible with alcohol, ether, and many other organic solvents	106.16 g/mol	C_8H_{10}	Solvent	372 mM
	Cyclooctane	4.54	Solubility in water = 7.9mg/L	112.21 g/mol	C_8H_{16}	Cycloalkane	405 mM
	Butylbenzene	4.72	Insoluble in water. Miscible with alcohol, ether, benzene.	134.22g/mol	$C_{10}H_{14}$		352 mM
	1,2,4,5-Tetramethylbenzene	4.28	Insoluble in water; freely soluble in alcohol, ether, benzene	134.22g/mol	$C_{10}H_{14}$		319 mM
	Tetralin	4.14	Insoluble in water; miscible with ethanol, butanol, acetone, benzene, ether, chloroform, petr ether, Decalin; soluble in methanol: 50.6% w/w	132.22g/mol	$C_{10}H_{12}$	Degreasing agent. Solvent for naphthalene	313mM
	Naphthalene	3.49	Soluble in water at approximately 30 mg/L	142.20 g/mol	$C_{11}H_{10}$		245 mM
	Decane	varies (<33%)	Immiscible in water	142.29 g/mol	$CH_3(CH_2)_8CH_3$	Alkane	< 2.3 M
	Undecane	varies (<33%)	Partially soluble in methanol, diethyl ether	156.31 g/mol	$C_{11}H_{24}$	Alkane	< 2.1 M
	Dodecane	varies (<33%)	Easily soluble in diethyl ether, acetone. Insoluble in cold water. Very soluble in Ethyl alcohol, Chloroform, Carbon Tetrachloride. Solubility in water: $3.7x10^{-3}$mg/L at 25°C	170.34 g/mol	$C_{12}H_{26}$	Alkane	< 1.94 M
	Tetradecane	varies (<33%)	Solubility in water: $2.2x10^{-3}$mg/L at 25°C	198.39 g/mol	$C_{14}H_{30}$	Alkane	< 1.66 M
	Hexadecane	varies (<33%)	Insoluble in water	226.45 g/mol	$C_{16}H_{34}$	Alkane	< 1.46 M
	Eicosane	varies (<33%)	Soluble in diethyl ether. Insoluble in cold water.	282.55 g/mol	$C_{20}H_{42}$	Alkane	< 1.17 M

Figure 2 - JP-8 solubility in human plasma. JP-8 was added to human plasma in glass tubes at the volumes indicated and mixed overnight at 37°C on a slowly rotating wheel. The resulting mixture was centrifuged (4,000 x g for 10 min) at room temp. An expanded image of the interface in the 5 ml JP-8/ 5 ml plasma mixture is also shown.

In vivo models are available to investigate the effects of JP-8 exposure but these models are limited in their capacity to uncover the molecular mechanisms of the direct effect of JP-8 exposure on cells. Alternatively, in vitro toxicological studies can be used to investigate toxicity mechanisms and inform in vivo studies where necessary. We have developed an in vitro exposure model for cell cultures using human plasma as a carrier for JP-8 that is physiologically relevant because all routes of in vivo exposure (dermal,

respiratory or oral) would result in the toxic component(s) of JP-8 carried and distributed in plasma.

We initially found that JP-8 was toxic when added directly to cells in culture (Figure 3).

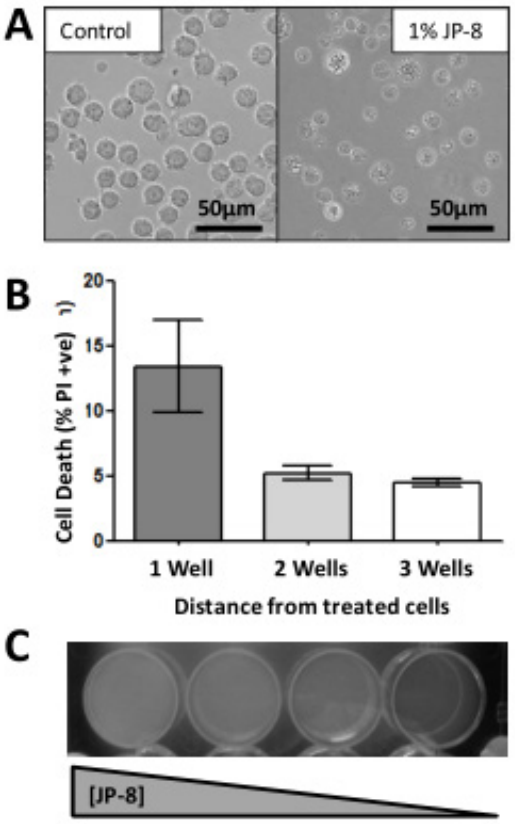


Figure 3. U937 cells were cultured in 24 well plates and treated with JP-8 added directly to culture media A) Cells treated with JP-8 were imaged after 24 h or B) the cells in adjacent wells were assayed by propidium iodide uptake at 24 hrs using flow cytometry. C) Images of the lid of a 24 well plate of U937 cells treated with different concentrations of JP-8.

However this strategy was unsatisfactory because we were unable to add low enough volumes of JP-8 to cells to obtain a toxicity profile. Further, the majority of JP-8 floated on the top of the media, making it difficult to

ascertain the level of JP-8 that entered the media. We also observed that cells in wells adjacent to the treated cultures also died .

Volatile components of JP-8 may therefore transfer from well to well and adversely affect results obtained from experiments performed in this way. The lids of culture dishes became opaque if JP-8 was added directly to the cultured cells and this interaction of JP-8 with the plastic may have an indirect, non-physiological, effect on the viability of cells in vitro. Finally, strong smelling vapours were released into the cell incubator when JP-8 was added directly to cultured cells, which posed potential risk to the investigator. Direct addition of JP-8 to cultured cells was therefore not deal for investigating the effects of JP-8 on cells.

In the current study we demonstrated that human plasma exposed to JP-8 can be used to deliver toxic components of JP-8 to cells in culture.

Method for preparing exposed plasma

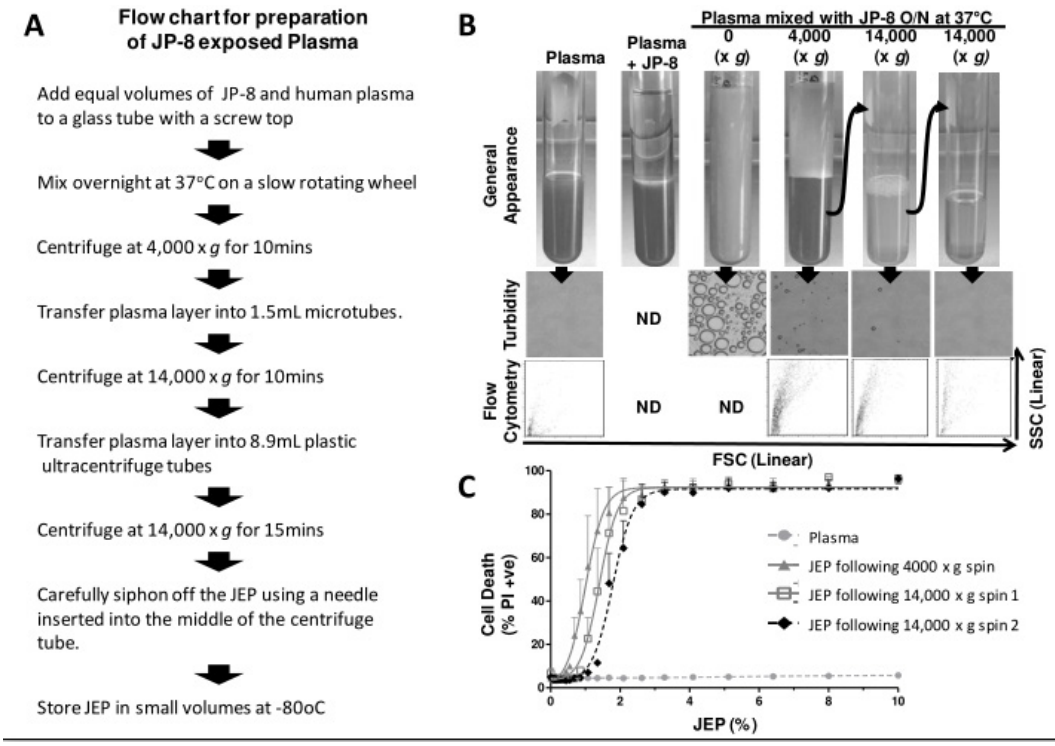


Figure 4 - Preparation and toxicity of JP-8-exposed plasma (JEP). JP-8 was mixed with an equal volume of human plasma followed by constant rotation overnight at 37°C. A) Serial centrifugations using the conditions indicated in the flow chart were used to isolate and clarify the plasma layer. B) The general appearance, turbidity of the mixture and presence of microscopic droplets was determined by imaging and flow cytometry at each step in the process. C) The toxicity of JP-8/plasma mixture following the 4,000 x g and each 14,000 x g centrifugation steps was determined by flow cytometric measurement of propidium iodide uptake in treated U937 cells at 24 h. Data presented is the average +/- SEM of three independent experiments.

The method we developed used glass tubes to mix the JP-8 solvents with plasma to overcome any interaction of the JP-8 solvents with the mixing vessel. The exposed plasma we produced dissolved completely when added to the culture medium and did not adversely affect viability of cells in neighbouring wells (not shown). The exposed plasma did not visibly affect the culture plastic when added to the cultured cells and did not emit strong smelling odours in the culture incubator.

Our method reduces method-related issues that may have affected previous studies where JP-8 was either added directly to cells in culture medium or delivered by non-physiological carriers (e.g ethanol), thereby presenting a more relevant, safer, quantifiable and reliable method of delivery. The data presented using this method is also consistent with previous research that showed components of JP-8 mixed with cell culture media were toxic to keratinocyte cultures and a haematopoietic cell line despite the insoluble nature of JP-8 in serum free culture media. Previous studies have employed ethanol as a delivery vehicle for JP-8. These studies found that JP-8 was toxic to cells at concentrations as low as 0.001% v/v. We also observed that JP-8 dissolved in ethanol was highly toxic to U937 cells with an LC50 of approximately 0.03% v/v, which was significantly higher than the LC50 for JEP (1.5%). This suggests that previous studies which used ethanol to deliver JP-8 may have over-estimated the toxicity of JP-8 by up to 15 fold.

It is important to note that the exact concentrations of JP-8 used in each study cannot be compared directly because of the manner in which the preparations were made (direct solubility vs exposure). This study also cannot be used to determine the concentration of JP-8 that is likely to be toxic in vivo or how much JP-8 is likely to enter the body because the

transport and delivery mechanisms through skin,. respiratory epithelium and gut mucosa are different. The different volatility of each of the components of JP-8 and chemicals used in aircraft maintenance may also lead to differences in exposure of cells. However it is likely the alveolar membrane is capable of delivering high levels of these chemicals into plasma given chemical similarities between JP-8 and chemicals used in aircraft maintenance with certain inhaled anaesthetic agents. It is interesting to speculate why JEP is toxic to cells.

In our study, the cells were treated with JEP in medium containing 10% FBS and L-glutamine, which contains all of the factors required for cell growth in culture. This suggests that the toxic effect of JEP is due to components of JP-8 that are transferred into human plasma rather than the effect of JP-8 stripping essential growth factors/nutrients from the plasma. This is supported by our biochemical comparisons which showed that the plasma layer from the JP-8/plasma mixtures did not differ significantly from control plasma in any plasma component detected.

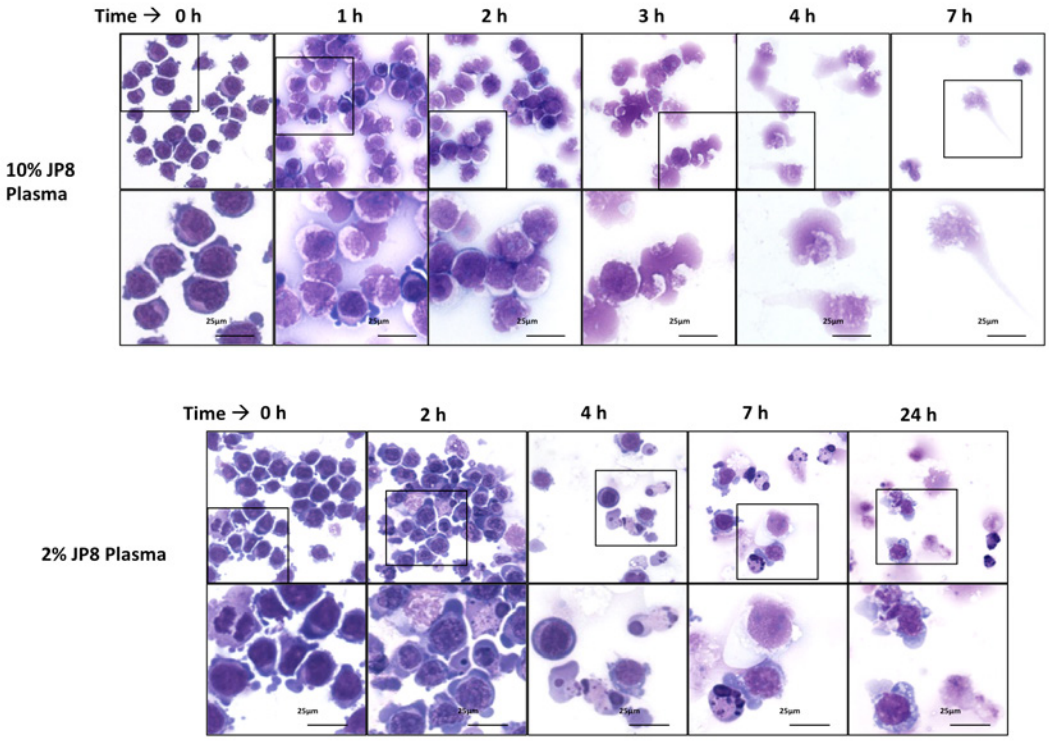


Figure 5. Showing cell death over time following exposure to different concentrations of plasma which had been saturated with JP8.

Differences between fuel components and solvents

The individual components of the JP8 and solvents were tested for preparation in the plasma model. Significant differences were found in the miscibility with plasma. These differences related to the polarity of the compound and the hydrophillic tendency to mix (dissolve) in plasma lipids.

This finding is significant in that it points to differences in the extent to which compounds would be taken up by plasma and distributed to internal organs.

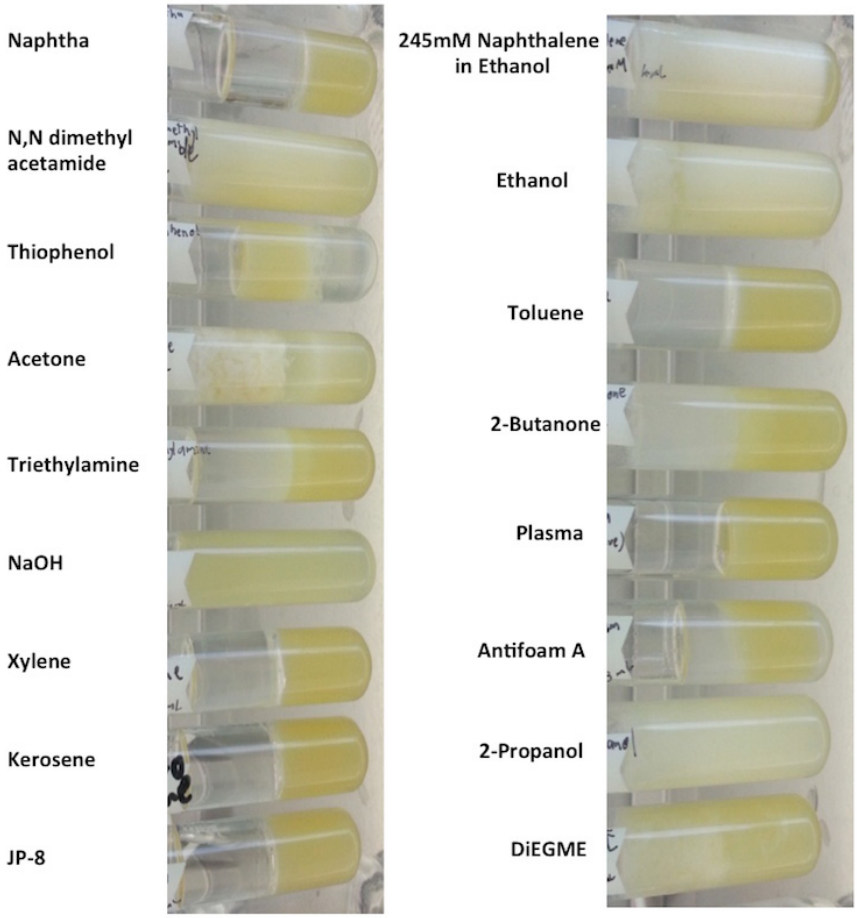


Figure 6. Interaction between DeSeal/ReSeal Compounds and human plasma. Some were completely mixable while others formed "soaps" with plasma lipids.

Mass Spectrometry Analysis of Plasma Binding

The exposed plasma was also subjected to analysis by mass spectrometry to quantify the extent to which individual compounds were bound to plasma and the mechanism by which they are transported in plasma .

Numerous hydrocarbons, naphthalene derivatives (and several unidentifiable peaks) were extracted from the JEP plasma and identified by mass spectrometry. However attempts at producing calibration data for titrated amounts of JP-8 components for absolute quantification were unsuccessful, possibly because the compounds were distributed in plasma proteins, lipoprotein complexes, and in lipids. However, for compounds identified and quantitated, the relative proportions of JP8 compounds bound to plasma components did not match the concentration distribution in the fuel. An example is shown in Figure 5 with significant enhancement of the relative quantity of napthalene in plasma compared to the quantity in JP8.

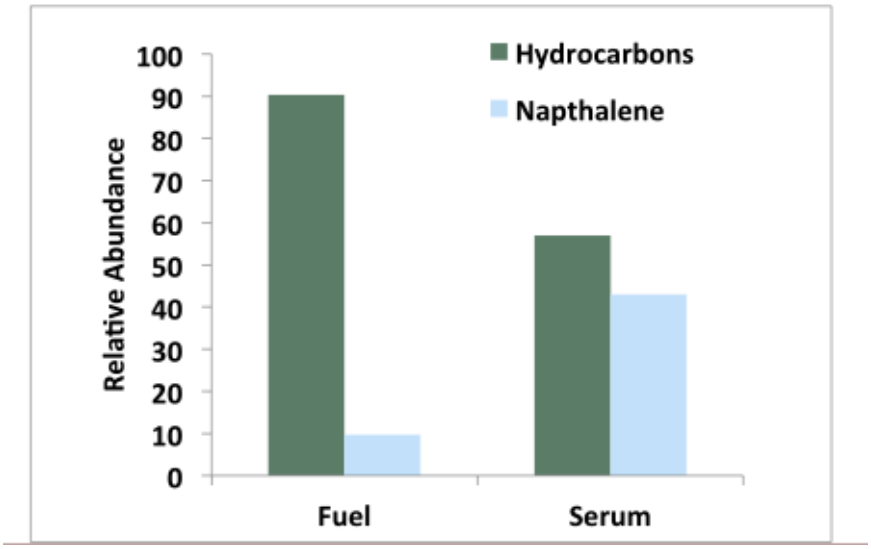


Figure 7. Comparison of the proportion of Napthalene in fuel compared to the proportion carried in human plasma showing a relative increase in the Napthalene carried and distributed to cells and tissues.

Investigation of other plasma components

In this study, we also investigated other potential physiological carriers of JP-8 such as commercial human serum, foetal bovine serum, and Albumin. In all cases we found that toxicity was reduced compared to exposed whole plasma. It is interesting to speculate on the compositional differences of these substances which may influence their capacity to transfer toxic components of JP-8 to cells. Serum is derived from plasma that has had clotting factors removed and perhaps these factors contribute to its reduced

capacity for toxic components of JP-8. Serum is often routinely collected in a fasting state, and lipid levels are proven to be lower. It is also possible that serum from different donors will have different capacity for these components or that serum that is stored for extended periods has reduced capacity for these compounds. Such differences may make some individuals more susceptible to JP-8 exposure than others.

This may also explain the differences in our results using human plasma and foetal bovine serum, which are different in their lipid and protein composition. Albumin is known to play a significant role in plasma transport. We used the human blood product Albumex 20 (200g/L albumin) to investigate whether the albumin in human plasma was a carrier for toxic JP-8 components but this was not the case as no cytotoxicity was observed in U937 cells treated with the Albumex 20/JP-8 mixture.

Therefore, other plasma components are responsible for the solubility of components of JP-8 of which lipids are an obvious candidate for further investigation as there are similarities between alkanes and plasma long chain fatty acids. Lipoprotein absorption could be considered as a possible transport mechanism of JP-8 in human plasma. Ultimately, using human plasma in preference to other sources will allow further investigations to determine how and why JP-8 affects cells and whether any of these effects can account for the pathology.

Scott et al., 2014. Figure 4

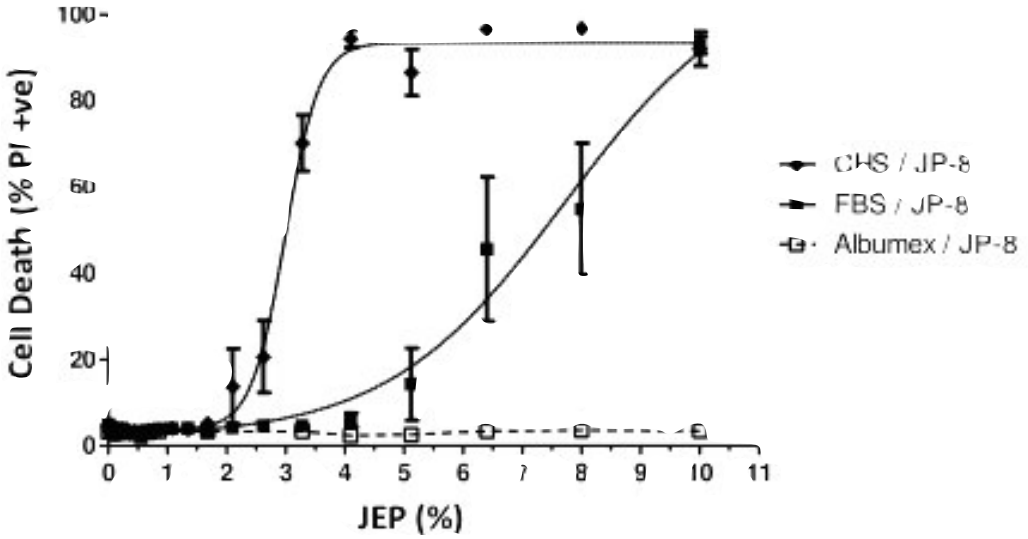


Figure 8. Cytotoxicity of JP-8 mixed with commercial human serum, foetal bovine serum or human albumin (Albumex 20). U937 cells were treated with commercial human serum (CHS), foetal bovine serum (FBS) or Albumex 20 (20% human albumin) that had been mixed with JP-8 and clarified using the same method as for human plasma. Cell death was measured at 24 h by propidium iodide uptake using flow cytometry. Data is the average +/- SEM of three independent experiments observed in personnel exposed to JP-8.

Response of various cell types to treatment with JP8 exposed plasma

The toxicity model was also tested against a variety of cultured cell types.

The cultured cell lines chosen were NK cells which are derivative of CD lymphocytes, REH cells which are a precursor B-cell, YT cells which are a transformed T cell, and RPMI which are a plasma cell myeloma derivative.

All cell types were susceptible.

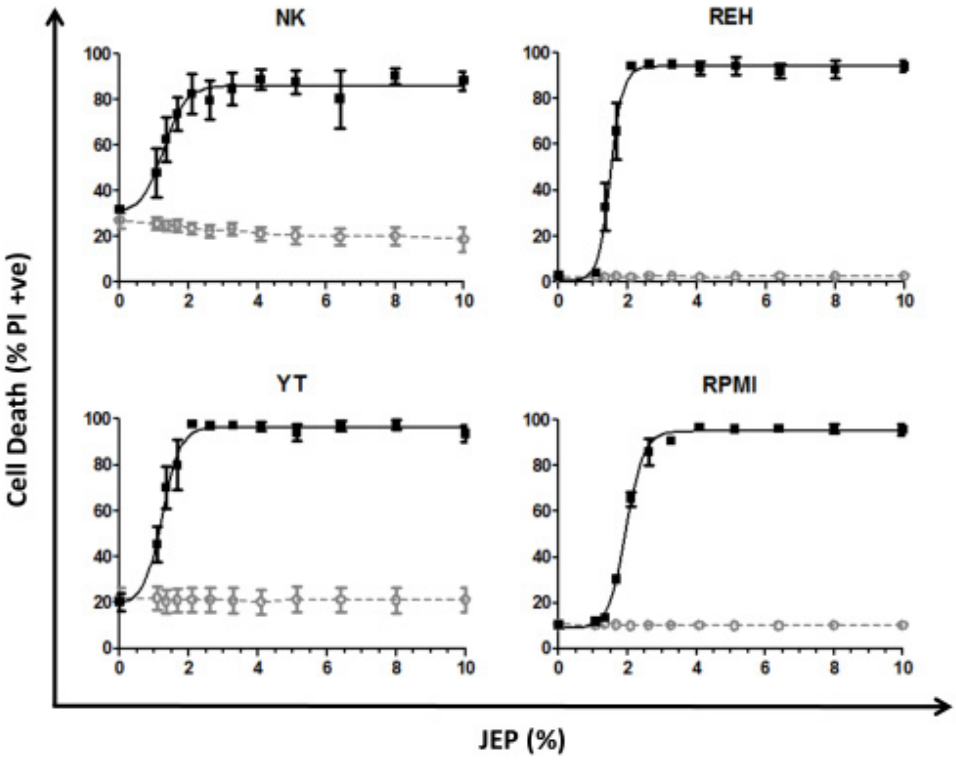


Figure 9. - JEP displays cytotoxicity for various human haematopoietic cells. Cultured cells were treated with various concentrations of JEP and cell death was determined by propidium iodide uptake at 24h using flow cytometry. Data is the average +/- SEM of three independent experiments.

Finally, it must be noted that toxicity is only one potential outcome of JP-8 exposure that could lead to disease. Our study provides a physiological model for studying various cellular effects of JP-8, including toxicity, which may help uncover the contribution of JP-8 exposure to a variety of pathologies and also raises the possibility of tracking and limiting exposure to these compounds in the future.

.

Determination of toxicity of individual components of JP8 and DeSeal/ ReSeal solvents

Background to the cytotoxicity assays.

This model (above) was then used in identification of toxic components of JP-8, which may have affected the exposed workers. Exposed plasma

preparations were made for each of the components of JP8 and solvents used in the DeSeal/ReSeal.

Overview of cell responses to injury

A cell initiates intracellular apoptotic signalling in response to a stress, which may bring about cell suicide. The binding of nuclear receptors by glucocorticoids, heat, radiation, nutrient deprivation,viral infection,hypoxia and increased intracellular calcium concentration, for example, by damage to the membrane, can all trigger the release of intracellular apoptotic signals by a damaged cell. A number of cellular components, such as poly ADP ribose polymerase, may also help regulate apoptosis.

Before the actual process of cell death is precipitated by enzymes, apoptotic signals must cause regulatory proteins to initiate the apoptosis pathway. This step allows apoptotic signals to cause cell death, or the process to be stopped, should the cell no longer need to die. Several proteins are involved, but two main methods of regulation have been identified: targeting mitochondria functionality, or directly transducing the signal via adaptor proteins to the apoptotic mechanisms. Another extrinsic pathway for initiation identified in several toxin studies is an increase in calcium concentration within a cell caused by drug activity, which also can cause apoptosis via a calcium binding protease calpain.

Mitochondrial regulation

The mitochondria are essential to multicellular life. Without them, a cell ceases to respire aerobically and quickly dies. This fact forms the basis for some apoptotic pathways. Apoptotic proteins that target mitochondria affect them in different ways. They may cause mitochondrial swelling through the formation of membrane pores, or they may increase the permeability of the mitochondrial membrane and cause apoptotic effectors to leak out.These are very closely related to intrinsic pathway, and tumors arise more frequently through intrinsic pathway than the extrinsic pathway because of sensitivity.There is also a growing body of evidence indicating that nitric

oxide is able to induce apoptosis by helping to dissipate the membrane potential of mitochondria and therefore make it more permeable.

Mitochondrial proteins known as SMACs (small mitochondria-derived activator of caspases) are released into the cytosol following an increase in permeability. SMAC binds to *inhibitor of apoptosis proteins* (IAPs) and deactivates them, preventing the IAPs from arresting the apoptotic process and therefore allowing apoptosis to proceed. IAP also normally suppresses the activity of a group of cysteine proteases called caspases, which carry out the degradation of the cell, therefore the actual degradation enzymes can be seen to be indirectly regulated by mitochondrial permeability.

Cytochrome c is also released from mitochondria due to formation of a channel, the mitochondrial apoptosis-induced channel (MAC), in the outer mitochondrial membrane, and serves a regulatory function as it precedes morphological change associated with apoptosis. Once cytochrome c is released it binds with Apoptotic protease activating factor - 1 (*Apaf-1*) and ATP, which then bind to *pro-caspase-9* to create a protein complex known as an apoptosome. The apoptosome cleaves the pro-caspase to its active form of caspase-9, which in turn activates the effector *caspase-3*.

MAC (not to be confused with the Membrane Attack Complex formed by complement activation, also commonly denoted as MAC), also called "Mitochondrial Outer Membrane Permeabilization Pore" is regulated by various proteins, such as those encoded by the mammalian Bcl-2 family of anti-apoptopic genes, the homologs of the *ced-9* gene found in *C. elegans*.*Bcl-2* proteins are able to promote or inhibit apoptosis by direct action on MAC/MOMPP. Bax and/or Bak form the pore, while Bcl-2, Bcl-xL or Mcl-1 inhibit its formation. Two theories of the direct initiation of apoptotic mechanisms in mammals have been suggested: the *TNF-induced* (tumour necrosis factor) model and the *Fas-Fas ligand-mediated* model, both involving receptors of the *TNF receptor* (TNFR) family coupled to extrinsic signals.

TNF Path

TNF is a cytokine produced mainly by activated macrophages, and is the major extrinsic mediator of apoptosis. Most cells in the human body have two receptors for TNF: *TNF-R1* and *TNF-R2*. The binding of TNF to *TNF-R1* has been shown to initiate the pathway that leads to caspase activation via the intermediate membrane proteins *TNF receptor-associated death domain* (TRADD) and *Fas-associated death domain protein* (FADD). cIAP1/2 can inhibit TNF-α signaling by binding to TRAF2. FLIP inhibits the activation of caspase-8. Binding of this receptor can also indirectly lead to the activation of transcription factors involved in cell survival and inflammatory responses. However, signalling through TNF-R1 might also induce apoptosis in a Caspase-independent manner. The link between TNF and apoptosis shows why an abnormal production of TNF plays a fundamental role in several human diseases, especially in autoimmune diseases.

Fas Path

The Fas receptor - First apoptosis signal(Fas) (also known as *Apo-1* or *CD95*) binds the Fas ligand (FasL), a transmembrane protein part of the TNF family. [23] The interaction between Fas and FasL results in the formation of the *death-inducing signaling complex* (DISC), which contains the FADD, caspase-8 and caspase-10. In some types of cells (type I), processed caspase-8 directly activates other members of the caspase family, and triggers the execution of apoptosis of the cell. In other types of cells (type II), the *Fas*-DISC starts a feedback loop that spirals into increasing release of proapoptotic factors from mitochondria and the amplified activation of caspase-8.

Common components

Following *TNF-R1* and *Fas* activation in mammalian cells a balance between proapoptotic (BAX, BID, BAK, or BAD) and anti-apoptotic (*Bcl-Xl* and *Bcl-2*) members of the *Bcl-2* family is established. This balance is the proportion of proapoptotic homodimers that form in the outer-membrane of the mitochondrion. The proapoptotic homodimers are required to make the mitochondrial membrane permeable for the release of caspase activators such as cytochrome c and SMAC. Control of proapoptotic proteins under normal cell conditions of nonapoptotic cells is incompletely understood, but in general, Bax or Bak are activated by the activation of BH3-only proteins, part of the Bcl-2 family.

Caspases

Caspases play the central role in the transduction of DR apoptotic signals. Caspases are proteins that are highly conserved, cysteine-dependent aspartate-specific proteases. There are two types of caspases: initiator caspases, caspase and effector caspases, caspase. The activation of initiator caspases requires binding to specific oligomeric adaptor protein. Effector caspases are then activated by these active initiator caspases through proteolytic cleavage. The active effector caspases then proteolytically degrade a host of intracellular proteins to carry out the cell death program.

Caspase-independent apoptotic pathway

There also exists a caspase-independent apoptotic pathway that is mediated by AIF (apoptosis-inducing factor).

Execution

Many pathways and signals lead to apoptosis, but there is only one mechanism that actually causes the death of a cell. After a cell receives stimulus, it undergoes organised degradation of cellular organelles by activated proteolytic caspases. A cell undergoing apoptosis shows a characteristic morphology:-

- Cell shrinkage and rounding are shown because of the breakdown of the proteinaceous cytoskeleton by caspases.
- The cytoplasm appears dense, and the organelles appear tightly packed.
- Chromatin undergoes condensation into compact patches against the nuclear envelope (also known as the perinuclear envelope) in a process known as pyknosis, a hallmark of apoptosis.
- The nuclear envelope becomes discontinuous and the DNA inside it is fragmented in a process referred to as karyorrhexis. The nucleus breaks into several discrete *chromatin bodies* or *nucleosomal units* due to the degradation of DNA.
- The cell membrane shows irregular buds known as blebs.
- The cell breaks apart into several vesicles called *apoptotic bodies*, which are then phagocytosed.

Apoptosis progresses quickly and its products are quickly removed, making it difficult to detect or visualise. During karyorrhexis, endonuclease activation leaves short DNA fragments, regularly spaced in size.

Cells growing in culture medium were exposed to increasing concentrations of the individual compounds. The cellular response was monitored using two different markers of cell death.

The Annexin V assay

In molecular biology, an **annexin A5 affinity assay** is a test to quantify the number of cells undergoing apoptosis. The assay uses the protein annexin A5 to tag apoptotic and dead cells, and the numbers are then counted using either flow cytometry or a fluorescence microscope.

The annexin a5 protein binds to apoptotic cells in a calcium-dependent manner using phosphatidylserine-containing membrane surfaces that are usually present only on the inner leaflet of the membrane.

Apoptosis is a form of programmed cell death that is used by the body to remove unwanted, damaged, or senescent cells from tissues. Removal of apoptotic cells is carried out via phagocytosis by white blood cells such as macrophages and dendritic cells. Phagocytic white blood cells recognise apoptotic cells by their exposure of negatively charged phospholipids (phosphatidylserine) on the cell surface.In normal cells, the negative phospholipids reside on the inner side of the cellular membrane and the outer surface of the membrane is occupied by phospholipids, which do not have a charge. After a cell has entered apoptosis, the negatively charged phospholipids are transported to the outer cell surface by a hypothetical protein known as scramblase. Phagocytic white blood cells express a receptor that can detect the negatively charged phospholipids. After detection the apoptotic cells are removed.

Detection of cell death with annexin A5

Healthy individual apoptotic cells are rapidly removed by phagocytes. However, in pathological processes, the removal of apoptotic cells may be delayed or even absent. Dying cells in tissue can be detected with annexin A5. Labeling of annexin A5 with fluorescent or radioactive molecules makes it possible to detect binding of labeled annexin A5 to the cell surface of apoptotic cells. After binding to the phospholipid surface, annexin A5 assembles into a trimeric cluster. This trimer consists of three annexin A5 molecules that are bound to each other via non-covalent protein-protein interactions. The formation of annexin A5 trimers results in the formation of a two-dimensional crystal lattice on the phospholipid membrane. This clustering of annexin A5 on the membrane greatly increases the intensity of annexin A5 when labeled with a fluorescent or radioactive probe. Two-dimensional crystal formation is believed to cause internalisation of annexin A5 through a novel process of endocytosis if it occurs on cells that are in the early phase of executing cell death. Internalisation amplifies additionally the intensity of the annexin A5 stained cell.

Detection of cell membrane damage with Propidium iodide

Propidium iodide (PI) staining based on cell membrane damage is a sensitive and quantitative method to test the cytotoxicity of typical lipophilic compounds. Cell membrane damage that resulted in increased PI uptake may be due to variations in the phospholipid and protein content of the membrane, which are affected by interactions between the lipophilic compounds and components of the cell membrane.

Results of cytotoxicity assays for JP8 components and Deseal/Reseal solvents

The relative toxicity of individual components when exposed to cell was classified according to the amount required to produce a cell response. A highly toxic response was defined by cell death when exposed plasma was mixed in cell culture medium at <5%. Medium toxicity was defined as toxic when delivered at between 5-10%. Low toxicity was defined as no appreciable toxicity when delivered at >10%.

Compound	Relative Toxicity
Acetone	Low
2 Butanone	Low
Antifoam A	Low
Ethanol	Low
Toluene	Low
Xylene	Medium
Ethyl Acetate	Low
Propylene Glycol Monomethyl Ether	Medium
2-Propanol	Medium
DiEGME	High
Kerosene	High
Napthalene	Medium
N,N Dimethyl acetimide	High
Sodium Hydroxide	High
Naptha	High
Thiophenol	High

Table 2. Relative toxicity of compounds of JP8 and Deseal/Reseal solvents

Cell Survival

The cell survival on exposure to individual components are shown in Figures below:-

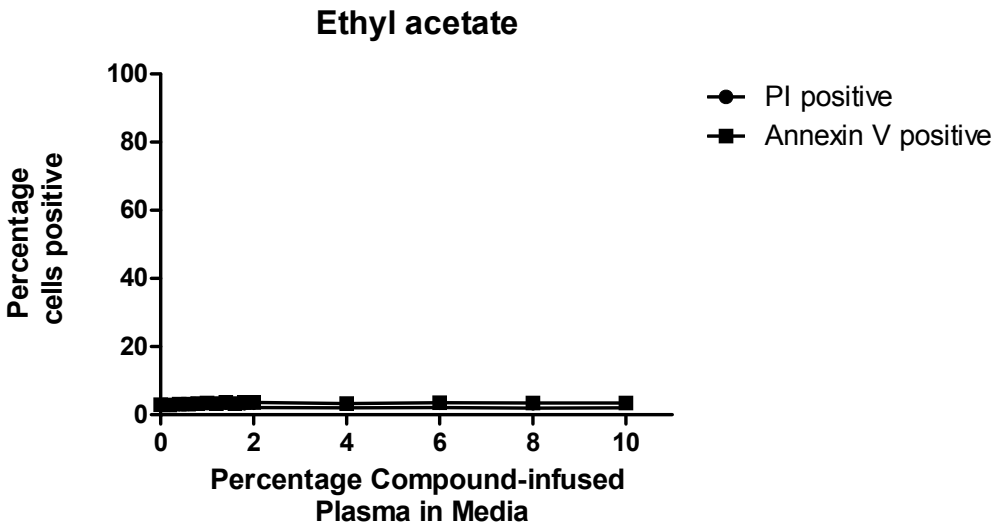


Figure 10.Ethyl acetate did not demonstrate toxicity

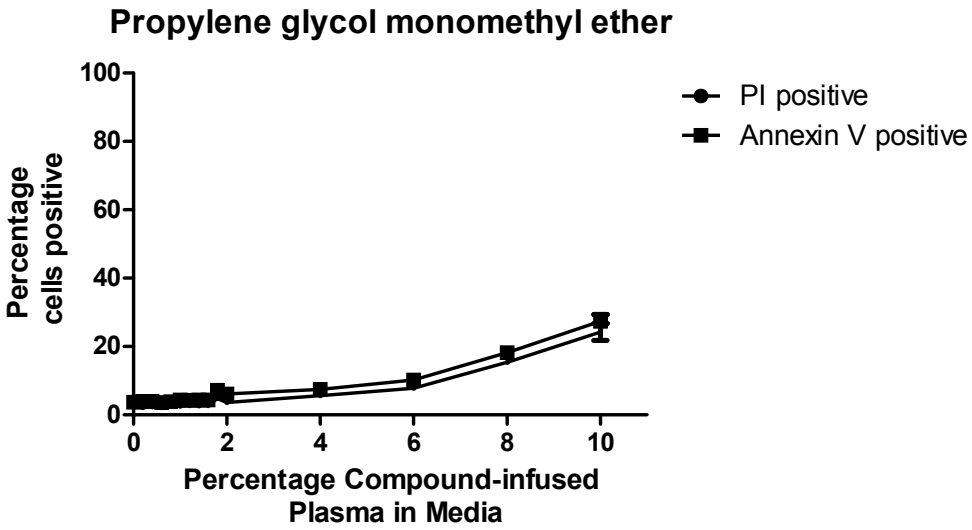


Figure 11. Propylene glycol with medium toxicity

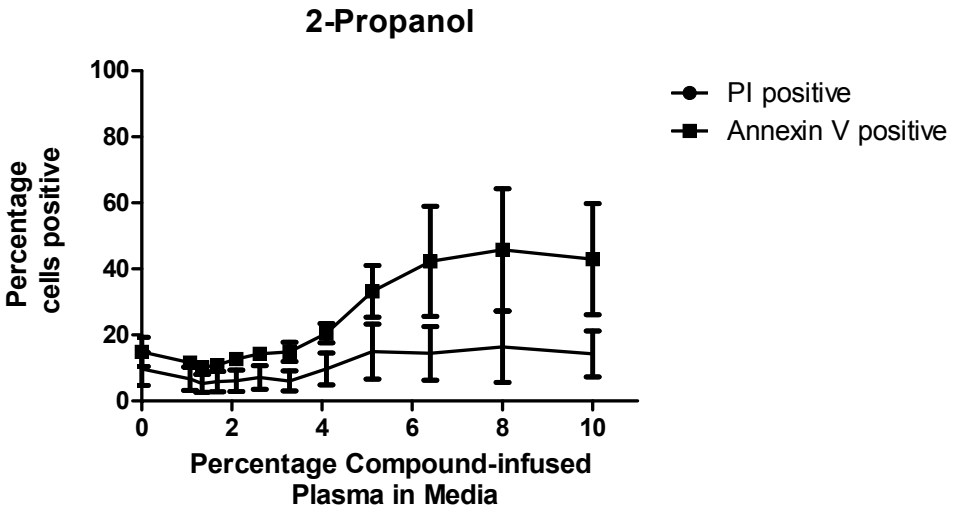


Figure 12. Propanolol demonstrated a different mechanism of cell injury with the apototic response mostly through the caspase pathway.

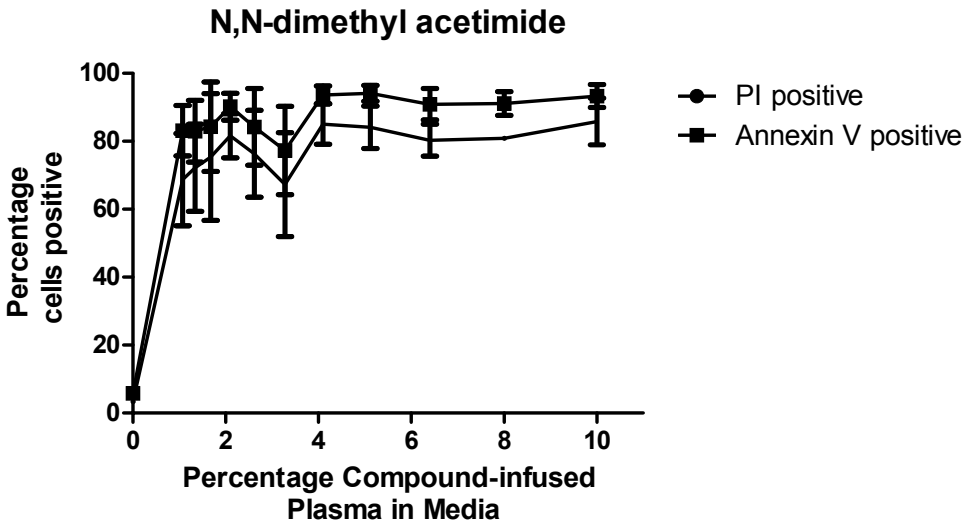


Figure 13. N,N - dimethyl acetamide showed high toxicity

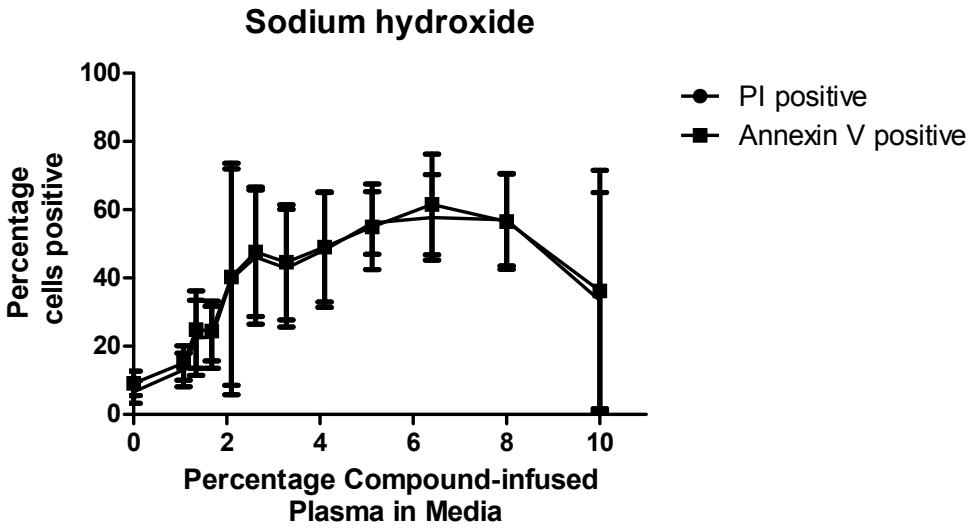


Figure 14. Sodium hydroxide demonstrated high toxicity

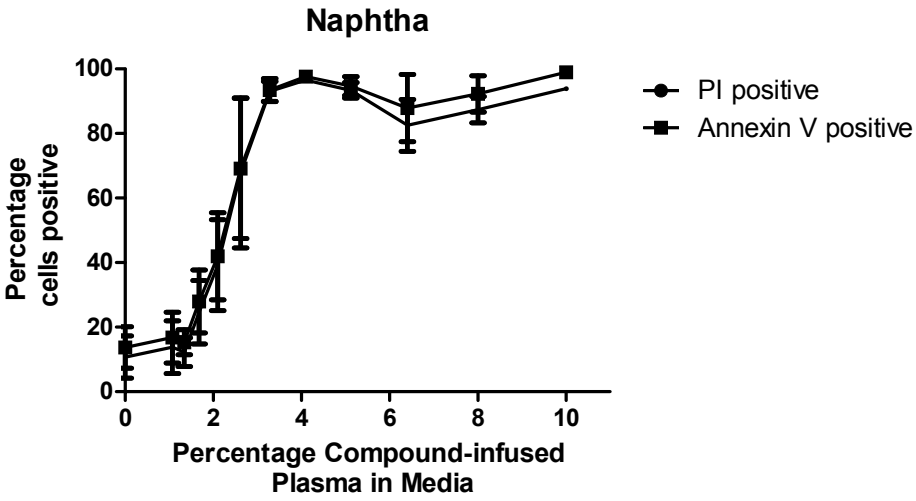


Figure 15. Naptha was highly toxic

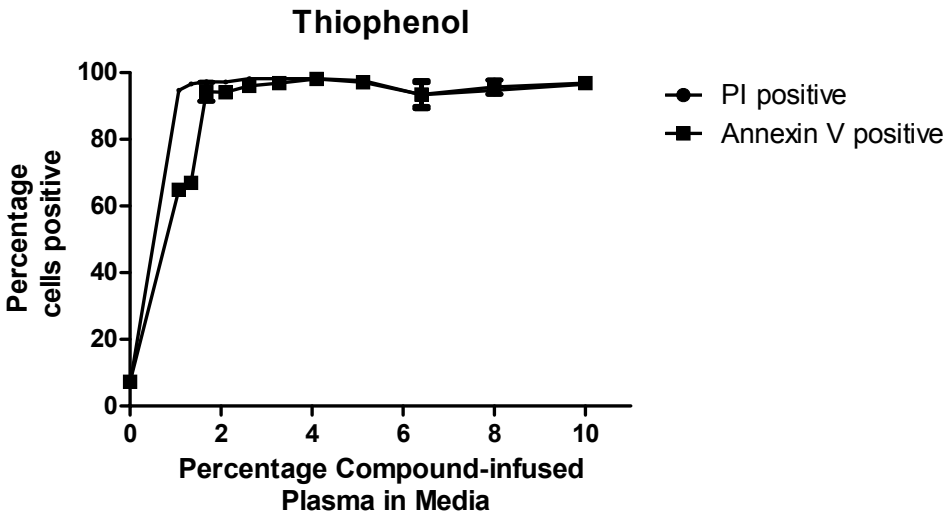


Figure 16. Thiphenol was highly toxic

Conclusions from cell exposure studies.

The compounds Kerosene, Naptha, Thiopenol, N,N Dimethyl acetimide demonstrated significant cytotoxicity when added to the cell culture media. Kerosene is a (variable) mixture of alkanes ranging in size from octane to hexadecane. The data are summarised as a kerosene mixture as each individual alkane demonstrated similar response. The cell death markers Annexin V and PI suggest that the cell death mostly occurred by disruption to the cell membrane initiating an apoptotic response. Earlier work by this group which implicated mitochondrial involvement in jet fuel toxicity through proteomic involvement may be explained by the role of mitochondria in the response of the injured cell through apoptosis.

Effects of exposure of cell gene expression

The effects of exposure to JP8 components and Deseal/Reseal solvents on cell biology were investigated changes to cell gene expression and to cell proteins. Results are summarised in the tables below:-

Table 3. Genes that were that had a five fold change in expression

Genes with expression changed >5 fold	Function of gene
Upregulated	
SUM01	Regulates sumoylation. BRCA1 is modified by SUM01 in response to genotoxic stress
ZNF417	Regulator of transcription
NOP16	Up-regulated in breast cancer. Regulates apoptosis via p581PK
TMEM167A	Transcriptional repressor. Elicits spontaneous immune responses in cancer patients
RLN3	Regulation of endocrine and autocrine/ paracrine hormones. Member of the insulin gene superfamily
Downregulated	
PRY	Similar to tyrosine phosphatase, unknown function
NEDD8	Regulates degradation of DNA damage response proteins following ubiquitination
STAP1	Regulates TEC protein kinase activity, which regulates cytokine receptor activity
OR4D1	Olfactory receptor
DNM1L	Dynamin GTPase. Establishes mitochondrial morphology. Mouse knockout has decreased neuritis and defective synapse formation
GK	Glycerol kinase. Involved in lipid metabolism

Table 4. The functions of groups of genes altered by exposure

Cell Survival/Death	NOP16 DYNLL1 FAM58A
Cell Metabolism	ISCA2 ENSA RLN3 GK TPTE
Cell Structure	DNM1L SMU1 HIST1H2BJ
Cell cycle/DNA damage	SUM01 SSX NEDD8 FAM58A
Cell Transcription/Translation	ZNF417 ZNF675 NEDD8 RPS4Y2 EIF2S2 RPL4

Table 5. Physiological role of clusters of genes involved in the response

Haematological	STAP1
Breast	NOP16
Olfactory	TASR20 OR4D1 OR5AU1 OR4N1 OR7C2
Metabolism	ISCA2 ENSA RNL3 GK TPTE
Cell Based Immunity	SSX2 DYNLL1 SLP1 CD1B
Neuronal	DYNLL1 SYP DNM1L

Table 6. Clusters of genes with known pathological associations identified in the cell response following exposure.

Cancer	NOP16 SSX2 IMP3 Sumo1
Immune suppression	SSX2 DYNLL1 SLP1 STAP1

Gene housekeeping	Sumo1 SSX2 EIF2S2 RPL4 FAM58 NEDD8 ZNF417/675 RPS4Y2

Table 7. Proteins most altered in cells treated with JP-8/Serum

Protein	Function
Up-regulated	
40S Ribosomal protein	RNA processing
Programmed cell death protein 5	May be involved in cell death
14-3-3 gamma	Signal transduction
Complement component 1Q	Antibody responses
Galactin 1	Binds complex carbohydrates
ZNF469	Transcription/translation
Down-regulated	
ATP synthase subunit delta	Similar to tyrosine phosphatase, unknown function
CD44	Receptor for hyaluronic acid
Nucleosomal assembly protein 1	Chromatin formation
Progesterone receptor component 1	Progesterone receptor
Acidic leucine rich nuclear – Phosphoprotein 32	Signal transduction

Table 8. Cell biology functions affected based on proteins most altered by JP-8/Serum

Cell Survival/Death
Cell Metabolism
Cell Structure
Cell cycle/DNA damage
Cell Transcription/Translation

Table 9. Pathological systems affected based on proteins most altered by JP-8/Serum

Pathology	Gene clusters identified
Cancer	14-3-3 Galectin 1 Mastermind like protein 2 Bcl-9
Myopathy	Galectin 1 Superoxide dismutase Transporter of the inner membrane Cytochrome oxidase subunit 5A
Cell Based Immunity	C1Q

CHAPTER 3

Study of Exposed Workers

3. Study of Exposed Workers

3.1 Overview

Genetic damage occurring as a result of jet fuel toxicity

The onset of pathology many years after exposure that was not detected shortly after exposure raises the possibility of epigenetic changes induced by the exposure as well at mitochondrial or nuclear gene mutations.

Genomic imprinting is an epigenetic process which can be studied using microarray investigations to assess gene expression profiles and methylation status (using two different microarray approaches). This process involves DNA methylation and histone modifications that result in alterations to gene expression. Such patterns of gene expression occur without any fundamental mutations to the DNA sequence pertaining to the gene of interest. That is, the imprint is via alteration to the gene control rather than the sequence. This is a dynamic process which can be erased and than re-established.

Mitochondrial DNA damage

In two case study investigations based on post-mortem data from Air Force maintenance staff that had worked on the Deseal/Reseal project, it was revealed that each carried a mutation in the mitochondrial DNA. Mitochondria are akin to mini power plants within a cell that supply cellular energy through the production of adenosine tri-phosphate (ATP). Mitochondria also control the programmed death of injured cells through a process called apoptosis. The novelty of this organelle is that it has its own genome that allows scientists to investigate changes in mitochondrial genes as well as nuclear genes. The two individuals who were the subject of this pilot investigation had succumbed rapidly to severe neurodegenerative disease with a temporal relationship to jet fuel and solvent exposure on the Deseal/Reseal project. Taken together, it suggests that there is a link between jet fuel/ solvent exposure and increased prevalence of morbidity and mortality in individuals with inherited mitochondrial defects.

Further research demonstrated that the levels of a protein referred to as TOM40 (important to mitochondrial repair and replication), were significantly

elevated in affected Air Force veterans that had worked in the Deseal/Reseal project. Such results indicate that the mitochondria had been exposed to a severe degree of sustained insult. Using a proteomic approach capable of assessing changes in a large number of biomarkers, it was observed that five proteins were consistently increased in all experimental samples.[9] Although not conclusive, there is sound reasoning and justification based on pre-clinical research and pilot data to investigate further in a large genomic and proteomic project using sophisticated bioinformatics methods, to definitively identify the relationship between the severity of jet fuel/solvent exposure and presentation of clinical phenotypes.

3.2 Study Plan

This study was conducted by Mater Research Institute (UQ) in collaboration with the Omics Laboratory within Mater Pathology over three years. Air Force veterans who worked on the De-seal/Re-seal project between 1973 and 1999 and/ or had direct or indirect exposure to F111 jet fuel were recruited. In addition, healthy controls matched for age and gender were recruited from within the general community.

An assessment of health outcomes was undertaken using a standardised General Health Assessment score. Study participants and their substitute decision maker (if applicable), were asked to attend a single visit for the purpose of blood collection at a local centre.

At the screening/study visit, Air Force veterans and healthy controls were enrolled into this trial if they had fulfilled all of the inclusion criteria and none of the exclusion criteria, and informed consent was complete. Clinical data was ascertained for the presence of any of the following conditions: Haematological malignancy, myelodysplasia, leukaemia, lymphoma, anaemia, immune deficiency, myopathy, cardiomyopathy, dementia, pychoses, any non-haematological neoplasia (benign and malignant, recurrent infection, bleeding disorders. This data was obtained both by interview of the subject.

Exposure data was obtained for each individual noting the duties which they undertook and the duration. Individuals were assigned to one of 3 groups based on the exposure history. These groups relate to dose of exposure and NOT to the Tier structure. If it was not possible to classify the exposure dose, then individuals were assumed to have been exposed to at the higher level. *(This essentially would minimise the significance of any changes seen supporting the Null Hypothesis*).

1. Definite high exposure by those who worked inside the fuel tanks.
2. Significant contact such as by dosing of skin, accidental ingestion, or mixing of the sealant
3. Lesser contact in the general area such as collection of rags or cleaning of the area.

A unique identification number assigned to participants at the clinic visit was recorded in the study database.

Primary endpoint assessment were determined by changes in blood count, genomic (gene expression and copy number) and proteomic variables using resources at the Mater Pathology Omics Laboratory. All serum, EDTA blood and PAXgene samples sent to this laboratory contained the participant's unique identification number, the participant's date of birth, and collection date. Analysis was performed using Affymetrix GeneChip analysis and LC MALDI TOF mass spectrometry. Masked results of genomic and proteomic laboratory reports were sent directly to the Principal Investigator and the Bioinformatition on completion of the test. Additional secondary endpoints were evaluated using outcomes from bioinformatic analysis of the genomic and proteomic results in-conjunction with medical, mental health and neurocognitive outcomes to establish risk determinants for prediction of future genotypic, phenotypic and clinical pathology.

Exploratory objectives were evaluated by toxicology and changes in signalling responses to biological stimuli determined from experimental procedures carried out at the MRI(UQ). Experimental data and respective analysis were collated and provided to the principal investigator. Human cell

lines were exposed to individual chemical mixtures and then subjected to genomic and proteomic analyses to identify individual agents or combination of agents able to produce cellular dysfunction.

Health Assessment Scores

General Health Questionnaire (GHQ30)

The purpose of the GHQ30 is to screen for medical disorders and non-psychotic psychiatric disorders. It is suitable for candidates 18+ years of age. The GHQ 30 provides a rapid and accurate assessment of general psychiatric status. It is a self-administered questionnaire that focuses on two major areas: the inability to carry out normal functions and the appearance of distress – to assess well-being in a person.

An individual's responses to Goldberg's 30-item General Health Questionnaire are usually represented as a single score which provides a measure of the number of psychiatric symptoms reported. No account is taken of the nature of the symptoms. Factor analyses of the GHQ-30 indicate an impressive degree of consistency of the factor structure, and the identification of five distinct factors corresponding to anxiety, feelings of incompetence, depression, difficulty in coping, and social dysfunction.

Molecular Investigations and Data Collection

Proteomics

A single SST tube was collected in order to conduct LC-MALDI TOF and LC-Electrospray MS investigations into the serum proteome. Additionally, the cellular proteome of leucocytes (derived from EDTA blood) was evaluated by 2 different approaches: LC-MALDI and LC-Electrospray mass spectroscopy. The mass spectroscopy platforms at the Omics laboratory include the Bruker Ultraflex III currently and MicroTOF Q-II.

Genomics

The aim of the genomics component of the JFESS was to use capture a comprehensive profile of the genome and transcriptome from a readily available tissue from men operating in the Deseal / Reseal programs and

aged-matched non fuel and solvent exposed male controls. Peripheral blood mononucleated cells (PBMCs) were chosen for the ease of collection and as one of the hypothesised tissues to have been exposed to any potential toxins during the Deseal / Reseal programs. To establish not only a comprehensive, robust and reproducible profile for each participant that can be compared between individuals the nucleic acid content of the collected PBMCs requires temporal fixation. This is achieved using collection tubes spiked with fixation solution specific for the fixation of DNA or RNA. The methodology chosen for extraction of nucleic acid can influence the range of DNA fragments or the relative ratio of RNA species isolated from a sample. One way to minimise variation in a dataset is to consistently perform extractions using the same protocol under the same conditions, accordingly extractions were conducted on an automated robot using magnetic bead purification to produce high quality nucleic acid samples.

The nuclear chromosome content encodes for almost the entire transcribed content within a cell, with an extremely minor but mitochondrial specific fraction coming from the mitochondrial chromosome. The regulation of gene expression can be achieved by a myriad of reversible and irreversible means including; gene deletion, gene duplication, gene silencing, expression of transcription factors, modification of DNA backbone (methylation, acetylation, phosphorylation and ubiquitination), histone modification and micro RNA interactions. To understand how an individual may respond to an external agent, an understanding of their capacity to express genes would give a basal prediction for their response, this can be ascertained by assessment of the number of gene deletions and duplications they carry. DNA samples were assayed by the Cytoscan HD (High Density) GeneChip® to describe how intact each participant's genome was, if any genome contained regions of DNA loss or duplication. The Cytoscan assay provides copy number information and homozygous / heterozygous status on the nuclear genome. The mitochondrial genome was also assayed and by virtue of its smaller size at 16569 basepairs it was possible to sequence the entire

genome to determine the sequence of every expressed mitochondrial gene and control regions.

The transcribed content of a cell is contained within the RNA population, itself a mixture of ribosomal, transfer, messenger and micro RNAs. The more dynamic populations of RNA are the messenger (mRNA) and micro (miRNA) RNAs and these were assayed to determine the current state of gene expression within the PMBCs. The Human Transcriptome 2.0 GeneChip® covers 67,528 genes, both coding and non-coding as well as probes spanning splice junctions to allow for the determination of differential exon usage. The miRNA 2.0 GeneChip® covers 1,105 pre-miRNA and their mature counterparts as well as 2,334 other mixed snoRNA, CDBox RNA, H/ACA Box RNA and scaRNAs. It is possible from the combination of these datasets to describe the genes and pathways active in the participants PBMCs and determine any differences compared to the age-matched non fuel and solvent exposed controls.

DNA extraction

DNA was isolated from PaxgeneDNA blood tubes using the QiaSymphony® robot (Qiagen) which uses magnetic bead purification to yield high-molecular, high quality DNA. PaxgeneDNA tubes were stored at -70oC from 24 hours post-collection till 24 hours prior to extraction. Tubes were thawed at room temperature for 24 hours followed by extraction of 200µl of whole blood using protocol Blood200v6 and eluted into 50µl Qiagen Elution Buffer. DNA was quantified by UV absorbance on a NanoDrop 8000 UV-Vis Spectrophotometer (Thermo Scientific) and stored at 4oC till required.

RNA extraction

RNA was isolated from PaxgeneRNA blood tubes using the QiaSymphony® robot (Qiagen). PaxgeneRNA tubes were stored at -70oC from 24 hours post-collection till 24 hours prior to extraction, where tubes were thawed at room temperature followed by centrifugation and washing of the PBMC pellet

twice in PBS. Cell pellets where then treated by protocol Paxgene_RNA_v5 and eluted into 50µl Qiagen Elution Buffer. RNA was quantified by NanoDrop 8000 UV-Vis Spectrophotometer (Thermo Scientific) and stored at -70oC till required.

RNA qualification

All total RNA samples were assessed for integrity using the Experion RNA StdSens Analysis Kit (Bio-Rad). Each sample was heated at 70oC for 2 minutes prior to rapid chilling on a 4oC cooling chamber (Diversified Biotech). One microliter of each sample was assayed alongside the supplied RNA ladder in the presence of loading buffer containing a marker allowing for normalisation of RNA migration time. A RNA Quality Index (RQI) score is generated for each sample from the Experion Data Analysis Tools, version 3.2. The RQI score ranges from 0 -10 and for this study only samples with an RQI above 5.7 were assayed further, median RQI was 8.8 +/- 0.7.

Human Transcriptome Array 2.0 Array

Total RNA was prepared for analysis on the Human Transcriptome Array 2.0 using the WT Expression Kit and the WT Terminal Labelling Kits from Ambion and Affymetrix respectively (ThermoFisher and Affymetrix). Manufactures instructions were followed with 100ng RNA used as input for all samples. In brief, RNA was used as a template for reverse transcription to create an RNA-DNA hybrid by incubation with First Stand Buffer/Enzyme mix in the presence of a primer containing a T7 promoter sequence, for 1 hour at 25°C, then 1 hour at 42°C, then at least 2 minutes at 4°C. Second Stand cDNA synthesis was performed, in the presence of RNase H, using DNA Polymerase I incubated at 1 hour at 16°C, 10 minutes at 65°C, then at least 2 minutes at 4°C. Next, cRNA is produced using T7 RNA Polymerase incubated at 40°C for 16 hours before purification of the cRNA and normalisation of sample yield to 455ng/µl in 22µl (10µg total). The cRNA is reverse transcribed in the presence of random primers and a nucleotide mix

where dTTP is underrepresented relative to the other three dNTPs, with the difference in ratio supplied by dUTP. Firstly cRNA is incubated with random primers only for 5 minutes at 70°C, 5 minutes at 25°C then at least 2 minutes at 4°C before addition of the 2nd-Cycle Master Mix and incubation for 10 minutes at 25°C, 90 minutes at 42°C, 10 minutes at 70°C, then at least 2 minutes at 4 °C. Next the cRNA is hydrolysed by RNaseH, to leave single stranded cDNA, by incubation for 45 minutes at 37°C, 5 minutes at 95°C then at least 2 minutes at 4°C before magnetic bead purification and quantification. Samples yielding 5.5µg cDNA were fragmented using a combination of uracil-DNA glycosylase and apurinic/apyrimidinic (AP) endonuclease by incubation for 1 hour at 37°C, 2 minutes at 93°C then at least 2 minutes at 4°C before end labelling using terminal deoxynucleotidyl transferase and DNA Labelling Reagent by incubation for 1 hour at 37°C, 10 minutes at 70°C then at least 2 minutes at 4°C. The labelled cDNA sample was prepared for hybridisation to the HTA 2.0 Array (Affymetrix, California, USA) by addition of 3.7µl Control Oligonucleotide B2 (3nM), 11µl 20X Eukaryotic Hybridization Controls, 110µl 2X Hybridization Mix, 15.4µl DMSO and 19.9µl nuclease-free water (GeneChip Eukaryotic Hyb Control Kit and GeneChip Hyb, Wash and Stain Kit (Affymetrix)). The hybridisation solution was incubated at 99oC for 5 minutes followed by 45oC for 5 minutes before 200µl was injected onto the array. Arrays were incubated for 16 hours at 45°C and 60 rpm in a Hybridization Oven 645 (Affymetrix) before washing and staining on a Fluidics Station 450 using protocol FS450_001 and reagents from the GeneChip Hyb, Wash and Stain Kit (Affymetrix). Arrays were scanned on a GeneChip® Scanner 3000 7G and CEL files generated automatically by the Affymetrix Command Console Software, version 3.2.

Micro RNA 2.0 Array

Total RNA was prepared for microRNA analysis using the FlashTag® Biotin HSR RNA Labelling Kit for Affymetrix® GeneChip® miRNA Arrays from Genisphere (Genisphere). Manufactures instructions were followed with 100ng RNA used as input for all samples. Samples were incubated with ATP

and Poly A Polymerase for 15 minutes at 37oC to add a polyA tail to the 3' end of all RNA before addition of the proprietary biotinalyted 3DNA® dendrimer by ligation using T4 DNA Ligase by incubation for 30 minutes at 25oC. The 23.5µl of labelled RNA was prepared for hybridisation to the miRNA 2.0 Array (Affymetrix) by addition of 15µl 27.5% Formamide (FlashTag kit), 50µl 2X Hybridization Mix and 10µl DMSO, both from GeneChip Hyb, Wash and Stain Kit (Affymetrix), 5µl 20X Eukaryotic Hybridization Controls and 1.7µl 3nM Control Oligonucleotide B2, both from the GeneChip Eukaryotic Hyb Control Kit (Affymetrix). The hybridisation cocktail was incubated at 99oC for 5 minutes followed by 45oC for 5 minutes before 100µl was injected onto the array. Arrays were incubated for 16 hours at 48°C and 60 rpm in a Hybridization Oven 645 (Affymetrix) before washing and staining on a Fluidics Station 450 using protocol FS450_003 and reagents from the GeneChip Hyb, Wash and Stain Kit (Affymetrix). Arrays were scanned on a GeneChip® Scanner 3000 7G and CEL files generated automatically by the Affymetrix Command Console Software, version 3.2.

Cytoscan HD Array

Genomic DNA was prepared for the Cytoscan HD Array using the CytoScan® HD Array Kit and Reagent Kit Bundle (Affymetrix) following manufactures instructions. In brief, 250ng of DNA at 50ng/µl was digested with NspI for 2 hours at 37oC then 20 minutes at 65oC to heat inactivate the restriction enzyme. Next, the NspI Adaptor sequence was attached to the digested DNA by incubation with T4 DNA Ligase at 16oC for 2 hours followed by 20 minutes at 70oC to heat inactivate the enzyme. The adaptor linked DNA fragments were diluted 1:4 and 10µl was amplified by PCR in four separate 100µl reactions containing: 39.5µl nuclease-free water (Affymetrix), 10µl 10X TITANIUM Taq PCR Buffer, 20µl GC Melt, 14µl dNTP Mixture (2.5mM each), 4.5µl PCR Primer 002 (Affymetrix) and 2.0µl 50X TITANIUM Taq DNA Polymerase (all reagents supplied by Clontech other than specified). The PCR was performed on an ABI 9700 with Gold Block in MAX mode with pre-incubation at 94°C for 3 minutes followed by 30 cycles of 94°C for 30

seconds, 60°C for 45 seconds, 68°C for 15 seconds finishing with 68°C for 7 minutes and hold at 4°C. The four PCR reactions were pooled and DNA isolated using supplied magnetic beads and eluted into 52µl of Elution Buffer from which 47µl was recovered for quantification and fragmentation. Only samples that yielded >3ug/µl as measured on a Nanodrop 2000 were subsequently processed. PCR samples were fragmented with 0.05U Fragmentation Reagent, 37°C for 35 minutes, 95°C for 15 minutes and held at 4°C before end labelling using terminal deoxynucleotidyl transferase and DNA Labelling Reagent by incubation at 37oC for 4 hours, heat inactivation at 95 °C for 15 minutes and hold at 4°C. The hybridisation solution was prepared using 70.5µl labelled DNA and 165µl Hyb Buffer Part 1, 15µl Hyb Buffer Part 2, 7.0µl Hyb Buffer Part 3, 1.0µl Hyb Buffer Part 4 and 2.0µl Oligo Control Reagent 0100 before incubation at 95oC for 10 minutes and held at 49oC prior to loading of 200µl onto the Cytoscan HD Array. Arrays were incubated for 16 hours at 50°C and 60 rpm in a Hybridization Oven 645 (Affymetrix) before washing and staining on a Fluidics Station 450 using protocol CytoScanHD_Array_450 and reagents from the CytoScan® HD Array Kit and Reagent Kit (Affymetrix). Arrays were scanned on a GeneChip® Scanner 3000 7G and CEL files generated automatically by the Affymetrix Command Console Software, version 3.2.

Mitochondrial Genome sequencing

The mitochondrial genome was first enriched by PCR from the total DNA sample then sequenced by massively parallel sequencing technology. The TaKaRa Long Range PCR kit (version 2.1,Clontech) was used to amplify the mitochondrial genome (mtDNA) in three separate PCR reactions producing amplicons of 3968, 5513 and 7814bp. Primer pairs used were: Mito1–Forward (ACA TAG CAC ATT ACA GTC AAA TCC CTT CTC GTC CC), Mito1-Reverse (TGA GAT TGT TTG GGC TAC TGC TCG CAG TGC), Mito2-Forward (TAC TCA ATC CTC TGA TCA GGG TGA GCA TCA AAC TC), Mito2-Reverse (GCT TGG ATT AAG GCG ACA GCG ATT TCT AGG ATA GT), Mito3-Forward (TCA TTT TTA TTG CCA CAA CTA ACC TCC TCG GAC TC) and Mito3-

Reverse (CGT GAT GTC TTA TTT AAG GGG AAC GTG TGG GCT AT). PCR reactions were performed in 25µl containing 50ng DNA, 14.75µl nuclease free water, 2.5µl 10X Buffer (containing 25mM MgCl2), 4µl dNTPs (2.5mM each), 1µl 10µM primer mix and 0.25µl TaKaRa LA Taq™ (5U/ul), with PCR cycled as pre-incubation at 94°C for 2 minutes followed by 30 cycles of 94°C for 15 seconds, 68°C for 8 minutes finishing with 68°C for 13 minutes and hold at 12°C. Amplicon yield was assessed by QuBit assay (Life Technologies) and equimolar amounts of the three PCRs pooled and purified using QIAquick PCR purification columns (Qiagen) and yield quantified on a Nanodrop 2000. Each pooled sample was then prepared for sequencing on a MiSeq instrument. Firstly samples were simultaneously tagged and fragmented using the Nextera XT transposome by incubation at 55°C for 5 minutes then held at 10°C till reaction termination by addition of neutralisation buffer. Next a limited cycle PCR was performed in the presence of Index 1 and Index 2 sequences, selected in combination such that up to 96 samples could be individually identified by their unique Index 1 and Index 2 sequence combination. PCR was performed using the Nextera PCR Master Mix under the following conditions; pre-incubation at 72°C for 3 minutes, 95°C for 30 seconds followed by 12 cycles of 95°C for 10 seconds, 55°C for 30 seconds, 72°C for 30 seconds finishing with 72°C for 5 minutes and hold at 10°C. PCRs were purified using AMPure XP magnetic beads, normalised using Library Normalisation Beads and all samples were pooled into a single tube ready for sequencing. Sequencing was performed according to manufacturer's instructions, using 600µl of pooled library and a 600 cycle version 3 kit (Illumina). MiSeq control software binned each sample according to the Index 1 / Index 2 pairing, produced paired forward and reverse FASTQ files that were used as input into Genomics Workbench v6.5 (CLC-Bio, Qiagen) where sequence variants were identified by mapping to NC_012920 genome (Homo sapiens mitochondrion, complete genome 16,569 bp circular DNA) using a Probabilistic Variant model to call variants and generated VCF v4.1 files.

Ranking of exposure

A modified exposure index based on the SHOAMP study was employed but simplified into 3 levels (High, Moderate, Low) . DSRS exposure was confirmed via a self-administered Exposure Questionnaire specifically developed for this study, with respondents indicating the specific DSRS program(s) they had been involved in , the tasks they performed, and for how long.

Data Quality Assurance

Data Handling and Storage

Clinical and experimental laboratory data was managed according to the standard procedures of the MRI(UQ). Copies of the completed CRFs were entered into the study database. Procedures for the treatment of missing, unused and spurious data was addressed in the detailed Statistical Analysis Plan (SAP).

Efficacy and Safety Variables

Primary Endpoint

Change in genomic and proteomic profiles between jet fuel/solvent exposed veterans and healthy controls.

Secondary Endpoint

Mortality and morbidity risk prediction based on genotypic, phenotypic and clinical outcomes measures.

Relationship between exposure indices and genomic/proteomic properties.

Statistical Considerations

Sample Size Determination

Statistical analyses have been developed to assess the hypothesis that genomic and proteomic profiles will be different when a group of Air Force Veterans who were exposed to jet fuel/solvents are compared with a healthy (age and gender-matched) control group.

Sample sizes were set up to allow 90% power to detect a 20% difference between gene and protein expression profiles, with an alpha value of 0.05, based on data from a similar study.

Statistical Methods

Statistical analyses were undertaken to assess the hypothesis that genomic and proteomic pathology will be different when a group of Air Force Veterans who were exposed to jet fuel/solvents are compared with a healthy (age and gender-matched) control group. Sample sizes were calculated to allow 90% power to detect a 20% difference between gene and protein expression profiles, with an alpha value of 0.05, based on observation from a similar study.

The statistical bioinformatic analyses was conducted by Emphron Informatics, and in consultation with the Principal Investigator and the Steering Committee. Statistical consultants were responsible for identifying and justifying the most appropriate statistical approach and producing a statistical report at study completion.

Analyses utilised an intent-to-analyse population and include all subjects with available data. No substitutions were made for missing data. Each data set, unless otherwise defined, was indexed on the participant initials and the unique identifying number.

Chi-square tests were performed on untransformed and on log-transformed data to demonstrate independence between genomic and proteomic outcomes for the healthy control group and the jet fuel/solvent (Air Force Veteran) exposed group. For the log transformed data, a 95% CI was given for the ratio of any two means. For the untransformed data, study group data was compared using the 95% CI to ascertain the difference between means.

Multivariate logistic regression analyses were conducted to determine relationships between molecular investigations, clinical data, and the exposure data in the jet fuel/solvent exposed group. It is possible that genetically determined susceptibility in a subset of individuals may be

identified. The effect may be independent of dose. The power of this observation cannot be determined until the data is obtained. An Odds Ratio will be presented to predict the risk of future genotypic, phenotypic and clinical pathology in the jet fuel/solvent exposed group.

Analyses utilised an intent-to-analyse population and included all subjects with available data. No substitutions were made for missing data. The analyses were defined in a detailed Statistical Analysis Plan (SAP) and the details of all analyses, listings and tabulations were defined. Any deviations from the SAP were clearly identified and justified in the final report.

All statistical analyses were performed using the statistical package, R. Each data set, unless otherwise defined were indexed on the participant initials and the unique identifying number.

The Intent-to-Analyse (ITA) Population comprises all participants who have either proteomic or genomic analyses conducted. All primary endpoints were analysed using the ITA population.

The Per-Protocol (PP) Population comprises all participants who have complete genomic and proteomic studies conducted for the purposes of this study. The Per-Protocol population were used for sensitivity analyses.

Where information was not available for data fields, it was marked as “Unknown” or “Not applicable”. All case report forms (CRF) were legible and only referred to the participant’s unique identifying number/code. The unique identifying number was documented in the CRF for the purpose of recording clinical data.

Bioinformatic Analysis of Proteomic and Genomic Outcomes

The quality of all RNA extracts were assessed prior to processing on gene chips. Gene expression estimates were developed using the RMA algorithm, applied to exon microarray data. XMAP annotation data were used to generate both gene level and exon level expression estimates. All chip data were screened using a variety of chip quality metrics (including but not limited to spike in concentrations, within-chip spatial correlations, and between chip correlations.

Proteomic analysis utilised proprietary software from Bruker (ProteinScape 1.3), capable of normalising spectra across multiple specimens and between runs; identifying statistically differentially expressed protein peaks in hundreds of samples; and subsequently identifying the nature of these individual proteins. Correlation of genes with proteins / peptides were greatly facilitated by our approach of running exon microarrays (rather than 3'UTR-based arrays), and utilised the algorithms described by Bitton et al, 2008.

Exploratory analysis of gene expression and proteomic data were based on principal components plots and hierarchical clustering.

In the absence of stratification, Chi-square tests were performed on untransformed and on log-transformed data to demonstrate independence between genomic and proteomic outcomes for the healthy control group and the jet fuel/solvent (Air Force Veteran) exposed group. In the presence of stratification, a Mantel Haenszel procedure was adopted.

All p values from association tests were adjusted to maintain a 5% false discovery rate.

A Hypergeometric Ontology enrichment analysis was used to identify molecular functions, biological processes, cellular components and pathways that are over-represented in the list of differentially expressed markers.

Multivariate logistic regression analyses were conducted to determine relationships between molecular investigations and clinical data in the jet fuel/solvent exposed group. Parsimonious models were fitted using stochastic variable elimination techniques based on a Laplacian prior. These techniques provide for computationally efficient selection from amongst massively multivariate putative predictors. Permutation tests of the overall association were used, and classification success will be assessed using multiple realisations of a *k*-fold cross validation scheme. In order to minimise selection bias, Cross validation was external to the marker selection process. An Odds Ratio was presented to predict the risk of future genotypic, phenotypic and clinical pathology in the jet fuel/solvent exposed group.

Review of Descriptive Medical Data

Full blood counts, and the Health assessment scores were summarised by level of exposure. Dichotomous variables were summarised by total number of observations, number and percentage of subjects with the characteristic. Continuous variables were summarised by mean, SD, median, minimum and maximum. Clinical laboratory values were presented by exposure group. Tables of selected laboratory parameters were provided where appropriate. Detailed subject listings of all laboratory data collected during the study were presented. Laboratory values outside normal limits were identified in the subject data listings with flags for high and low values.

Study population and recruitment

546 participants from a phenotype and exposure database (SHOAMP) that had already been created for the purpose of a previous study investigating the relationship between exposure to jet fuel/solvents in the Deseal/Reseal project and health outcomes responded to a request to participate in this study. This list was categorised by location and amount of exposure to jet fuel. 196 current and ex-Air Force personnel who worked either indirectly or directly on the Deseal/Reseal project between 1973 and 1999 and/ or had direct or indirect exposure to F111 jet fuel were approached to take part in the study in October 2012. 6 participants withdrew from the study, 10 could not be collected and 2 participants died during the recruitment phase of the study. 178 participants were given the opportunity to provide blood samples and complete health questionnaires between November 2012 and June 2013. In addition, 77 healthy controls that were matched for age and gender were also recruited from within the general community or current and ex-Air Force personnel were tested between June and December 2013.

Table 10. Study participants

Jet Fuel Exposure	Participants selected	Number Participants	Participation rate
High Exposure	96	84	87.5%
Moderate Exposure	50	45	90%
Low Exposure	50	46	92%
TOTAL	196	175	89.3%

Control population

In addition, 84 healthy controls that were matched for age and gender from within the general community or current and ex-Air Force personnel consented to be tested between June and December 2013. 7 controls withdrew from the study.

Table 11. Study controls

Jet Fuel Exposure	Participants approached	Number Participants	Participation rate
None	84	77	91.7%

Age matched controls

There were no statistical differences in the age distributions of controls or exposure groups. A Kruskal Wallis rank sum test (a non-parametric 1 way analysis of variance) obtained the result:-

$$\text{Kruskal-Wallis chi-squared} = 4.2433, \text{df} = 3, \text{p-value} = 0.2364$$

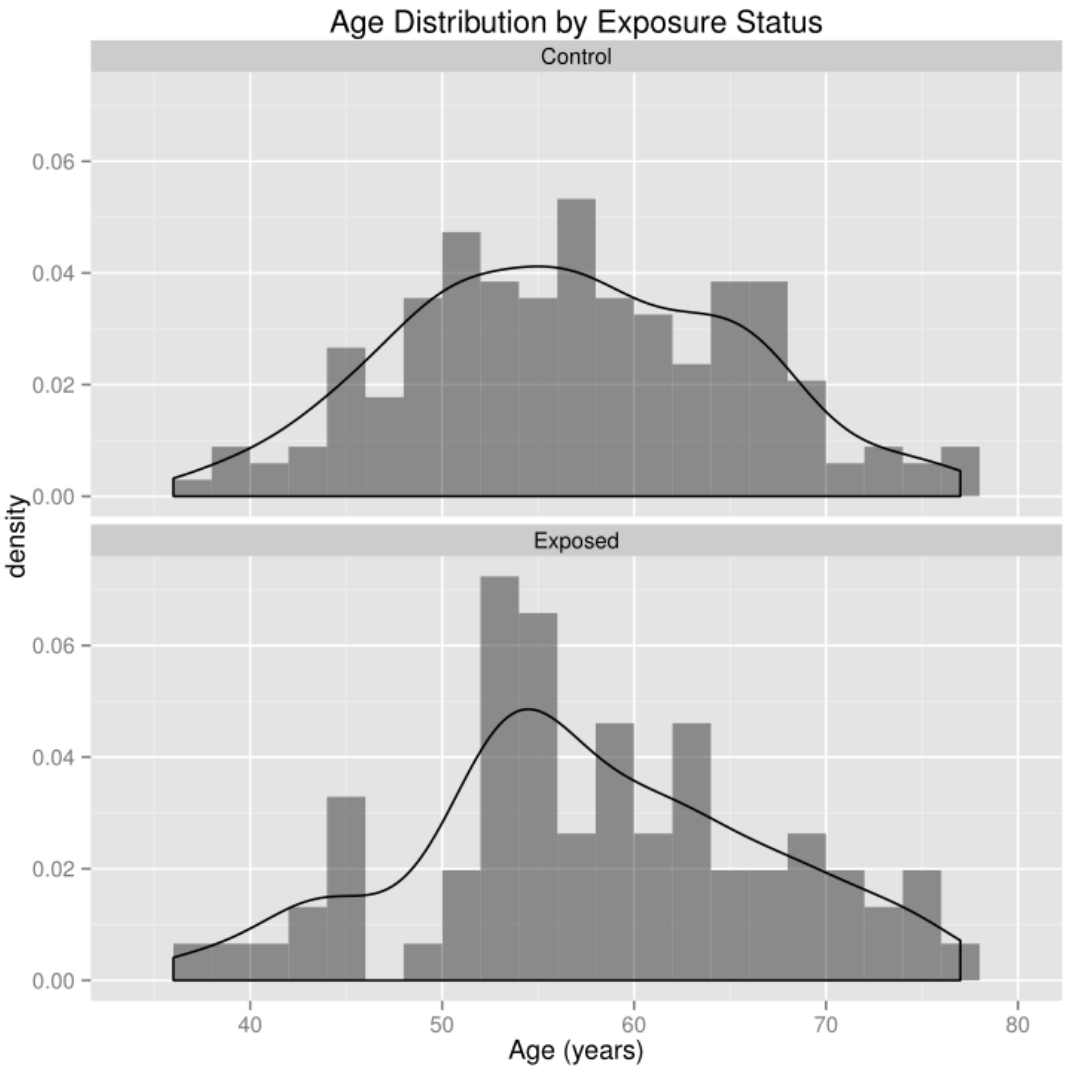


Figure 17. Comparison of age distribution of exposed workers and controls

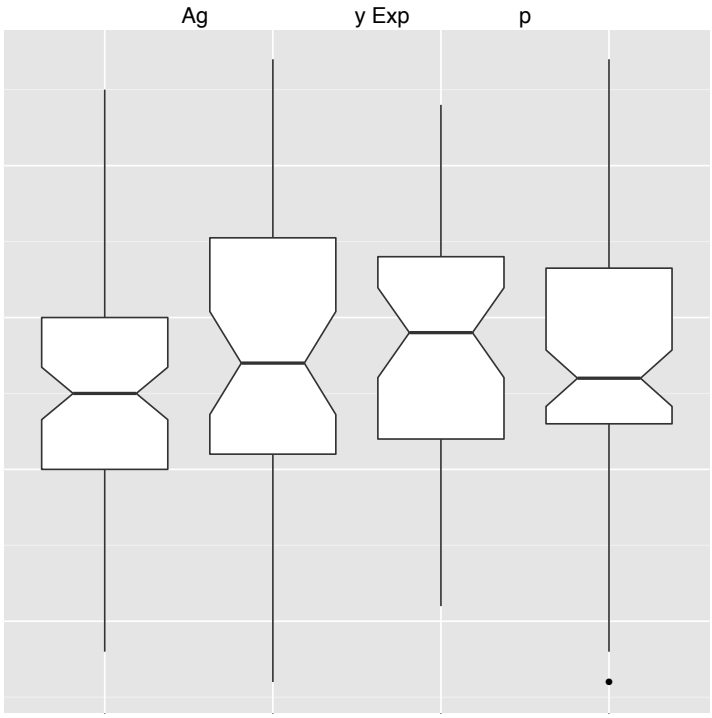


Figure 18. Comparison of age distribution by type of exposure (1 – High, 2 – Medium, 3 – Low) and Controls (4)

Ethics

Ethical approval for the project was obtained from Mater Health Services, the Australian Department of Defence and the Department of Veterans Affairs.
No ethical issues or complaints arose with participants or controls during the study.

3.3 Results of study of exposed workers

The study of exposed workers showed differences from the matched control group in health indices, and in some genomic studies. Details are presented below.

Health Assessment Scores

The Health assessment scores showed exposed workers to have a lower health rating than controls. There did not appear to be a decrease in the health scores (dose response) related to the degree of exposure. Workers with mild exposure had the same decrease in their health scores as those with high exposure.

Bioinformatic analysis of Genomic and Proteomic data and Health scores

Detailed test results and statistical analysis of the genomic, proteomic, and health assessments are include in Appendices 3-7.

The genetic studies of blood cells examined two types of changes in gene expression, and the presence of chromosomal changes and for mutations the mitochondrial DNA.

There were no chromosomal changes detected at a level of 50Kb using a high resolution SNP ARRAY

There was no change in the mitochondrial DNA mutation load between exposed workers and age matched controls. Mitochondrial DNA changes can accumulate with age.

There were no changes in the amount or type of protein coding mRNA expression, which is an index of cell activity. In disease states , these are usually tissue specific and may not appear in blood cells unless they are directly involved in the disease process.

There were small but significant and consistent changes in the expression of regulatory microRNAs that control activity of other genes. The regulatory functions of the altered genes have been linked to neurological changes and neurodegenerative disorders. It must be emphasised that interpretation of the

function of regulatory genes is an evolving science with much uncertainty at present. (The regulatory genes, which compose 98% of our genome, have a major role in human development, adaptation and response to disease. The function is only known for ~40% of these at present. Disease causing associations, with some early exceptions, are still unmapped).

Protein studies of plasma and blood cells

No significant changes were seen in the levels and types of protein expressed in the plasma and blood cells of exposed workers. A few small changes were seen consistently, but these did not reach a level that the researchers considered significant.

3.4 Discussion and Conclusions

Confounders and sensitivity

Lack of dose response

A dose response would have been expected but was not observed in the workers with different exposure histories. The unexpected similarity in the health scores and genomic studies within the exposed groups (low, medium, high) raises several hypotheses:-

The possibility of confounders

Confounders could include:-

An ascertainment bias whereby only those workers affected by any exposure volunteered to participate in the study.

An ascertainment bias whereby only those workers NOT affected by the exposure (ie. Survivors) volunteered to participate in the study.

The workers were stratified by their exposure to Deseal/Reseal materials. The effects seen may NOT be due to the Deseal/Reseal materials but to some other experience of the workers. The cellular studies suggest that exposure to fuel alone could be responsible.

It was not possible to examine other possible shared confounding events in the work careers or in the lifestyle of the personal. (eg. other occupational exposure not related to Deseal/Reseal such as medications, substance abuse, nutrition)

This study was conducted on individuals between 10 and 30 years after their exposure. If significant changes occurred at the time of exposure, normal cellular repair and selection mechanisms may have lessened the biological signal that could be observed in this study. The small but consistent changes observed suggest this possibility. Either the effect at the time was minimal but has persisted, or the effect was larger but has diminished overtime.

The cellular studies show that the compounds are mostly distributed by plasma lipids. The exposure to organs within the body would likely depend on the concentration of plasma lipids at the time of fuel exposure. Plasma lipids vary genetically between individuals, with lifestyle and alcohol intake, with composition of their diet, as well as the time after meals when the exposure occurred. The lack of a dose effect could be explained if workers in the lower exposure group had higher plasma lipids at the time of exposure. Individuals in the high exposure group worked within the fuel tanks and were selected because they were leaner and smaller, possibly protected to some extent by lower plasma lipids.

3.5 Significance of findings

The cellular findings, supported by other recently published genomic studies, indicate a definite toxicity from JP-8 to exposed cells. The components of JP8 tested are commonly found in most (aviation) fuels. The results indicate that there is a need for concern about exposure to fuels in general.

The study was not designed to determine the degree of occupational exposure necessary to produce cellular changes. However, the results show that cells grown in a nutrient containing as little as 5% exposed plasma are affected. In the body, blood cells have 100% exposure to plasma while other organs will have less exposure depending on the net blood flow and cellular membrane barriers. Organs such as brain, liver and bowel have very high blood flow. Cellular membranes generally have greater permeability to substances dissolved in lipids.

The study was also not designed to determine the most toxic routes of exposure (inhalation, ingestion, skin contact), but did demonstrate that fuel components can be distributed to organs through blood plasma, ie organs such as brain or liver, not directly exposed in the contact, may undergo secondary exposure. The implication is that all body systems must be considered in assessing/monitoring the health of exposed workers.

While the changes seen many years after exposure were small they were consistent within the group. The changes are most apparent in gene regulation and had some association to the health problems (eg, malignancy) identified in other studies. There were no chromosomal changes or mutations linked to the exposure. The genes changes seen can be described as Epigenetic, which is a mechanism of cellular adaptation to some environmental influence. Epigenetic changes are less clearly linked at the present time to disease. Epigenetic changes occur through a variety of cellular mechanisms and these were not investigated in this study. Some epigenetic changes can be transferred down through successive generations but to date have not been shown to cause birth defects or mutation in off-spring.

CHAPTER 4

Appendices, References, Glossary

Appendix 1
Ethical Statements

Ethics

Ethics approval was obtained from all Institutions involved in the study :-

Mater Health Services Human Research Ethics Committee **(***Ethics Approval MHS HREC.pdf)*

Australian Defence Human Research Ethics Committee , Directorate of Health Research Co-ordination, Joint Health Command Department of Defence) *ADHREC Approval 12 Nov 2012.pdf }*

DVA Human Research Ethics Committee, Australian Government, Department of Veterans' Affairs **(** *E010-027 - DVA Human Research Ethics Committee.pdf }*

Mater
Exceptional People. Exceptional Care.

MATER HEALTH SERVICES HUMAN RESEARCH ETHICS COMMITTEE

Prof Frank Bowling
c- A/Prof Nigel Waterhouse
Mater Medical Research Institute
Raymond Tce
South Brisbane QLD 4101

20th July 2011

Dear Prof Bowling

Re: Protocol *Ref №.1774A* Molecular Investigations into the basis of Disorders Manifested following Deseal/Reseal and Clinical Risk Prediction Associated with Jet Fuel Exposure Syndrome

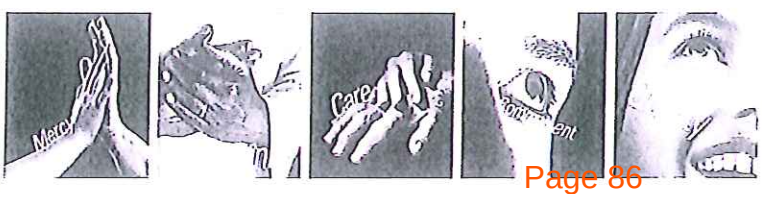
I write to advise that the Mater Health Services Human Research Ethics Committee considers the above study to meet the requirements of the *National Statement on Ethical Conduct in Human Research* (2007) and has granted ethical approval for your research proposal. Please accept our very best wishes for the success of this study. *In all future correspondence with the Committee please quote the Mater reference number.*

Documents reviewed and approved include:

- *Study Protocol Dated 22nd June 2011*
- *Participant Information and Consent Form Dated 22nd June 2011*
- *Australian Defence Human Research Ethics Committee Approval Letter*
- *Department of Veterans' Affairs Human Research Ethics Committee Approval Letter*

This approval is valid until **20th July 2014**. Please note the following conditions of approval.

- Any departure from the protocol detailed in your proposal must be reported immediately to the Committee.
- When you propose a change to an approved protocol, which you consider to be minor, you are required to submit a written request for approval to the Chairperson, through the Secretary. Such requests will be considered on a case by case basis and interim approval may be granted subject to ratification at the next meeting of the Committee.
- Where substantial changes to any approved protocol are proposed, you are required to submit a full, new proposal for consideration by the Human Research Ethics Committee.
- You are required to advise the Research Ethics Coordinator immediately of any complaints made, or expressions of concern raised, in relation to the study, or if any serious or unexpected adverse events occur.
- Under the *NHMRC National Statement on Ethical Conduct in Research Involving Humans*, research ethics committees are responsible for monitoring approved research to ensure continued compliance with ethical standards, and to determine the method of monitoring appropriate to each project. You are required to provide written reports on the progress of the approved project annually, the first report being due on **date** and finally on completion of the project. *(The Progress Report is located at http://www.mater.org.au/Home/Research/Human-Research-Ethics-Committee.aspx or can be accessed through the Mater Intranet, Applications, Research Register then under the project name or alternately can be emailed to you).* Please inform the Committee of publications, presentations at Conferences, education and quality improvement outcomes from this study. The Committee may also choose to conduct an interim audit of your research.



- Please be aware that all study procedures including follow up of participants and data analysis should be completed within the approval time frame or an extension should be requested.

You are reminded that this letter constitutes ethical approval only. You must not commence this research project until authorisation from the Research Governance Office has been obtained.

Please contact the Executive Director in the participating hospital/hospitals prior to commencing of the study. To access medical records, for the purpose of this study, please provide a copy of this approval letter to the Corporate Health Information Manager. I would also be grateful if you could confirm the date of commencement. (All correspondence should be directed to the Mater Research Ethics Coordinator.)

Yours sincerely

Dr Andrew Crowden
Chairperson
Mater Health Services Human Research Ethics Committee

Australian Government
Department of Veterans' Affairs

DVA HUMAN RESEARCH ETHICS COMMITTEE

Contact: HREC Coordinator
Telephone: (02) 6289 1129
Facsimile: (02) 6289 6296
E-mail: *ethics.committee@dva.gov.au*

Professor Francis Bowling
Mater Health Service
Level 2, Mater Children's Hospital
Raymond Terrace
SOUTH BRISBANE QLD 4101

Dear Professor Bowling

E10/027- Jet Fuel Exposure Syndrome Study (JFES Study)

Thank you for submitting a protocol change to the Department of Veterans' Affairs Human Research Ethics Committee (the Committee) for the above proposal. Consideration of the protocol changes were considered on 19 October 2012. The Committee agreed that they meet the requirements of the *National Statement on Ethical Conduct in Research Involving Humans* and therefore the Committee gave its approval for the change.

The Committee reserves the right at any time to seek further information about the research, noting that may affect continuation of its approval.

Yours sincerely

Kyleigh Heggie

Kyleigh Heggie
Director
Research Development and Coordination

13 November 2012

JOINT HEALTH COMMAND

ADHREC, CP3-6-036, Campbell Park Offices, PO Box 7911, Canberra BC ACT 2610

2010/1145828
ADHREC/OUT/2011/R13148748

Prof Frank Bowling
Prof Deon Venter
A/Prof Waterhouse
Miss Rachelle Warner

Dear Researchers

AUSTRALIAN DEFENCE HUMAN RESEARCH ETHICS COMMITTEE (ADHREC) COMMENTS ON PROTOCOL 605-10 - MOLECULAR INVESTIGATIONS INTO THE BASIS OF DISORDERS MANIFESTED FOLLOWING DESEAL/RESEAL AND CLINICAL RISK PREDICTION ASSOCIATED WITH JET FUEL EXPOSURE SYNDROME.

1. Thank you for submitting your research modification for consideration by ADHREC. Your modification request was considered by the committee out of session.

2. ADHREC has approved your modification noting:
   a. The new questionnaire issue was clarified. Some members felt it may not be as effective as the previously proposed questionnaire, but that it is a methodological, rather than an ethical issue, and the researchers' justification for the change was satisfactory.

   b. The consent form has now been adequately modified to communicate the destruction of the blood samples to participants.

3. Should you have any questions in relation to this matter please contact the ADHREC Secretariat. Contact details are provided below.

Yours sincerely,

Sarah Blackledge
Secretary
ADHREC

Tel (02) 6266 3807
Fax (02) 6266 2697
E-mail: ADHREC@defence.gov.au

12 November 2012

Appendix 2
Results of health assessments

Jet Fuel and Solvents Exposure Project

Health Questionnaire

Emphron Informatics Pty Ltd **www.emphron.com**

Report Number:	2012.020
Version Number:	1.0
Client:	Jet Fuel and Solvents Project
Date:	July 25, 2014
Author:	Mervyn Thomas
Email:	mervyn.thomas@emphron.com
Signature:	.....

1 Introduction

1.1 Background and Objectives

The Jet Fuel and Solvents Exposure study has been set up to investigate physiological and other differences between ex Air Force employees who were exposed to aviation fuel during their service. Subjects were stratified into four exposure groups, where Group 4 is a control group (representing no known exposure) and Group 1 is the highest level of exposure.

Data are available from several genomic and proteomic modalities, including: cytogenetics data, High Throughput mRNA gene expression data, micro RNA expression data, serum analysis by LCMS and full blood counts. This report is concerned with the analysis of a health questionnaire scale. The scale is generated as a sum of 30 items: each measured on a discrete scale. Higher values of the score represent poorer psychological states.

1.2 Available Data

Mitochondrial DNA variant data are available from 245 subjects. 84 of these subjects were in the high exposure group; 44 in the medium exposure group; 41 in the low exposure group and 76 in the control group.

2 Methods

A box and whisker plots[1] was generated, showing the difference in health scores between exposure groups. Between group differences were investigated using an analysis of variance. All pair-wise group comparisons were performed, using t tests based on the pooled estimate of residual variance from the analysis of variance

3 Results

The distribution of health scores by exposure group is shown in Figure 1. There is no suggestion of any differences between the exposure groups, but a clear difference between the exposure groups and the control group.

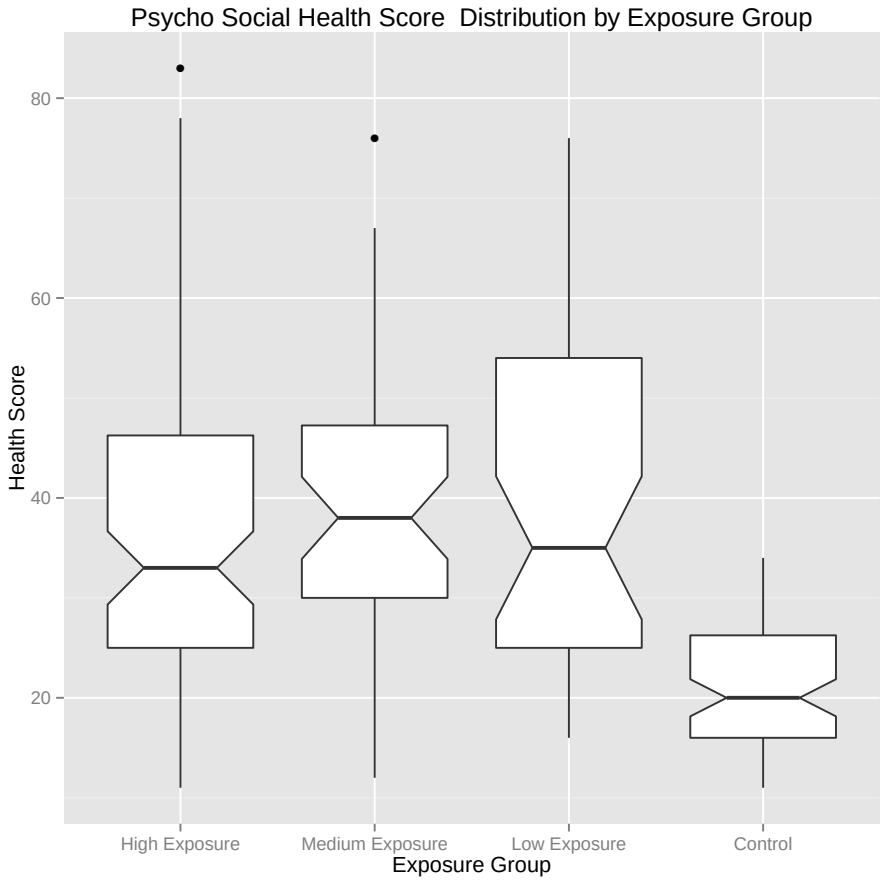


Figure 1: Distribution of Health Scores By Exposure Group

Table **??** shows the t statistics for the pair-wise comparisons between exposure groups. The control group is statistically significantly different from every other group. There are no statistically significant differences between the exposed groups.

	Exposure			
	High	Medium	Low	Control
High Exposure	0.00	1.18	1.32	-6.92
Medium Exposure	-1.18	0.00	0.15	-6.95
Low Exposure	-1.18	0.00	0.00	-6.96
Control	6.92	6.95	6.96	0.00

Table 1: t Statistics for Pairwise Comparisons Between Exposure Groups.The critical value for t at the 5% level is 1.97. The control group is statistically significantly different from every other group. There are no statistically significant differences between the exposed groups.

4 Discussion

There are clear statistically significant differences between the control group, and each of the exposed groups. The control group score is lower than the exposed groups: demonstrating that psychological health in the control groups is better than in the exposed groups.

There are demographic differences between the groups, and this result does not necessarily imply a causative relationship.

References

[1] J. M. Chambers, W. S. Cleveland, B. Kleiner, and P. Tukey. *Graphical Methods for Data Analysis*. Wadswoth and Brooks / Cole, 1983.

Appendix 3
Report from Emphron on Bioinformatic analysis of gene expression array data from leucocytes of exposed workers and controls.

ANALYSIS of HTA 1

Emphron Informatics Pty Ltd
Author: Mervyn Thomas
Date: 11 June 2014

Jet Fuel and Solvents Exposure Project

Analysis of High Throughput Arrays – HTA2.0

Emphron Informatics Pty Ltd **www.emphron.com**

Report Number: 2012.020
Version Number: 1.0

Client: Jet Fuel & Solvents Project
Date: July 27, 2014
Author: Mervyn Thomas
Email: mervyn.thomas@emphron.com

Signature: .....

Contents

List of Tables

List of Figures

1 Introduction

1.1 Background and Objectives

The Jet Fuel and Solvents Exposure study has been set up to investigate physiological and other differences between ex Air Force employees who were exposed to aviation fuel during their service. Subjects were stratified into four exposure groups, where Group 4 is a control group (representing no known exposure) and Group 1 is the highest level of exposure.

Data are also available from other genomic modalities (including cytogentics data and micro RNA data) and from proteomic sources. Only the results of gene expression from blood samples, estimated using the Affymetrix HTA2.0 chip are described in this report.

Each subject was also asked to complete a health questionnaire (Ref), and data from the psycho social scale of the questionnaire were investigated in conjunction with the gene expression.

1.2 Available Data

HTA gene chip data were available from 225 subjects. The break down of these subjects by exposure group and date at which the chips were scanned is shown in Table 1. We note that all group 4 chips were scanned on or after 16thOctober 2013. The comparison of Group 4 with the other groups is therefore partially confounded with date of scanning. This issue is addressed in the Methods section below.

Date	Exposure Group			
Scanned	1	2	3	4
2013-03-12	2	2	1	0
2013-07-19	1	1	1	0
2013-08-01	9	5	8	0
2013-08-02	9	4	3	0
2013-08-05	0	0	2	0
2013-08-07	0	1	0	0
2013-10-04	6	0	3	0
2013-10-11	11	3	9	0
2013-10-16	3	4	4	20
2013-11-06	14	5	1	8
2014-01-10	7	2	2	21
2014-03-19	8	4	4	11
2014-03-26	7	8	5	6

Table 1: Samples By Exposure Group and Date Scanned

2 Pre Processing of Gene Chip Data

Raw gene chip data were processed to give gene level estimates of expression using the RMA algorithm[6], as implemented in the Affymetrix Power Tools Suite[8]. Standard quality metrics were generated as recommended by Affymetrix[1].

Two of those quality metrics were selected for further examination, based on the author's experience of their performance on large data sets. The metrics considered were:

all+probesets_mad_residual_mean The mean of the absolute deviation of the residuals from the median. The RMA algorithm produces a model for probe level responses, based on the particular probe and the probe set to which it belongs. Residuals from this model represent departures from the model prediction for the given probe on the given chip. The mean absolute deviation for a chip provides a measure of how well the particular chip fits the RMA model: which was developed across all chips. A high mean absolute deviation of the residuals from the median suggests quality problems for the chip concerned.

all_probesets_rle_mean The mean absolute relative log expression (RLE) for all the probesets on a given chip. This metric calculates the $\log_2$ ratio of the intensity for each probeset to the median intensity for that probeset over all chips. The mean $\log_2$ ratio is calculated. This measure will vary with true biological variability between samples, but since in any one data set *most* genes will be unchanged, this metric is generally low for high quality data.

Chips were sorted by date and time of scanning, and exponentially weighted moving average control (EWMA) charts[10, 11] were generated for each metric. The EWMA chart is a powerful method of identifying when a process moved out of control. It is not intended to identify individual extreme values; the EWMA is more sensitive to runs of extreme values. It generally lags the onset of extreme values by a few observations.

Histograms were constructed for each metric, and robust Huber estimates[5] of the mean and standard deviation were used to generate mean $\pm$ 3 standard deviation ranges.

Figure 1 shows the EWMA plot for mean absolute deviation of residuals. There is a period towards the end of the sequence in which the process appears to be out of control. Figure 2 shows a histogram and kernel density estimate[12] of the distribution of this metric. The distribution is relatively well behaved, with a small number of extreme values. These extreme values occur during g the period highlighted on the EWMA chart.

Inspection of the data revealed that these extreme values were for five RNA samples obtained by manual extraction. The samples were dropped from down stream analysis.

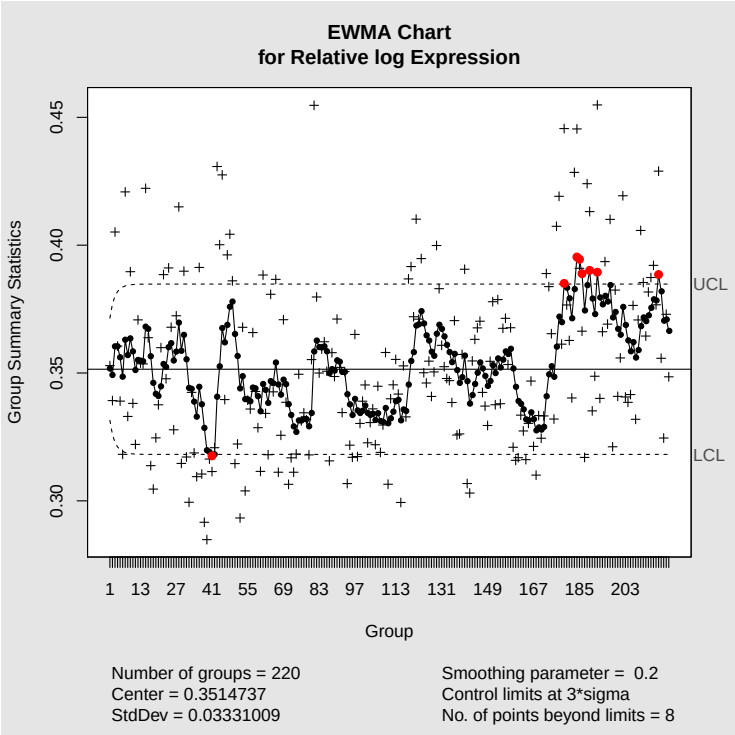


Figure 1: Exponentially Weighted Moving Average (EWMA) Control Chart for Mean Absolute Deviation of Residuals

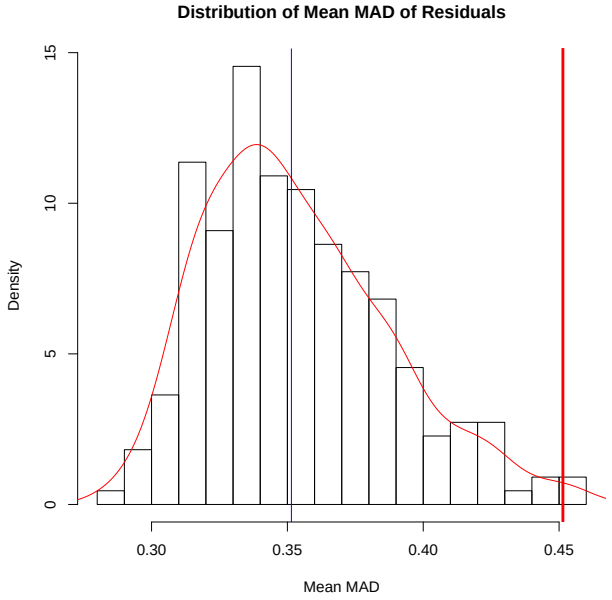


Figure 2: Distribution of Mean Absolute Deviation of Residuals

Figure 3 shows the EWMA control chart for the relative log expression metric. Again, there is a group of samples towards the end of the period in which the process appears to be out of control. The distribution of the metric is shown in Figure 4. Again, the five most extreme values correspond to the RNA samples obtained using manual extraction.

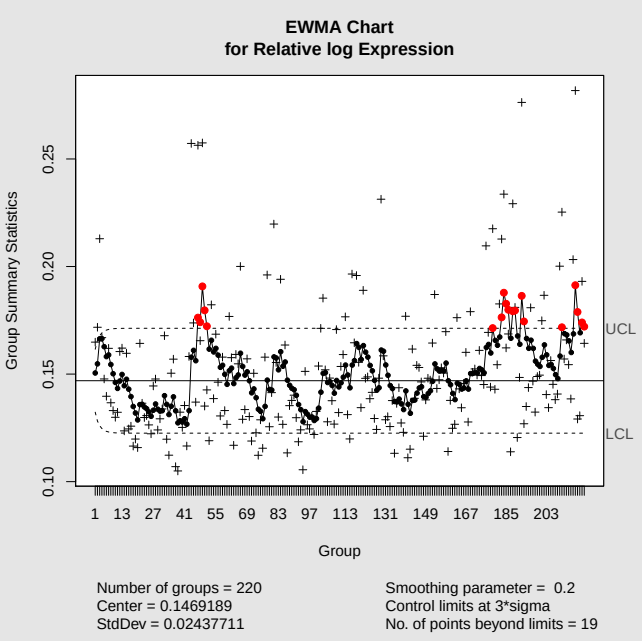


Figure 3: EWMA Chart for Relative Log Expression

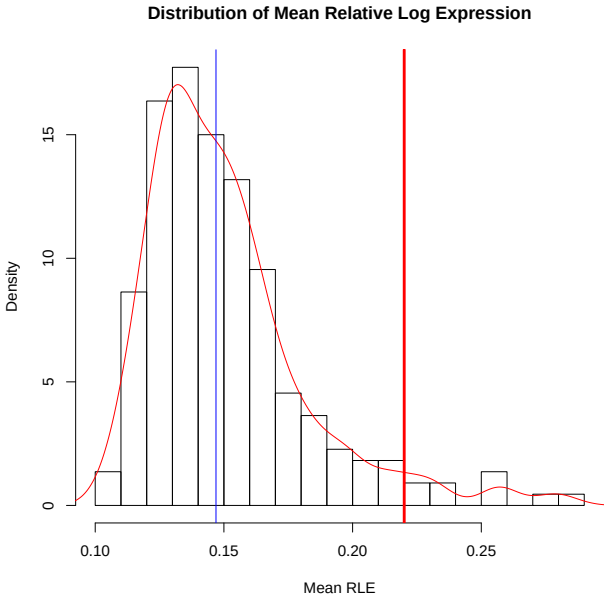


Figure 4: Distribution of Relative Log Expression

The quality metrics calculated for these data were very satisfactory. A small number of extreme values were associated with documented differences in chip processing. These chips have been excluded from further analysis.

Probe sets which had low overall expression (median across all chips $\log_2$ expression < 2.4) or low overall variability (median absolute deviation from median across all chips $\log_2$ expression < 0.064) were filtered out of all further analysis analysis.

3 Exploratory Analysis of Gene Expression

Principal components[7] of the gene expression matrix (represented as samples X genes) were calculated. The scree plot is shown in Figure 5. This plot shows the percentage of total variance explained fir the first 10 components. The plot shows a rapid decline in variance for the first four components, followed by a long tail. The first four components account for a total of 44% of the total variance.

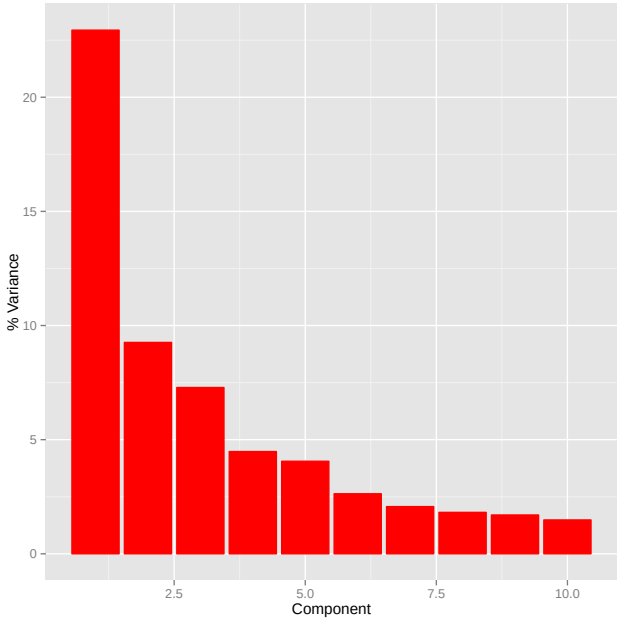


Figure 5: Scree Plot For Principal Components Showing Percent Variance Explained By the First 10 Components.

Figure 6 shows the disposition of data for each chip in the space of the first two principal components. Exposure group is colour coded. The plot reveals no evidence of structure, and no suggestion that the first two principal components are associated with exposure to jet fuel and solvents. Figure 7 shows similar results for components 3 and 4. Again there is no suggestion of any relationship with exposure group. The implication is that if there are differences between the Jet Fuel and Solvent exposure groups, they are not a major source of chip to chip variation.

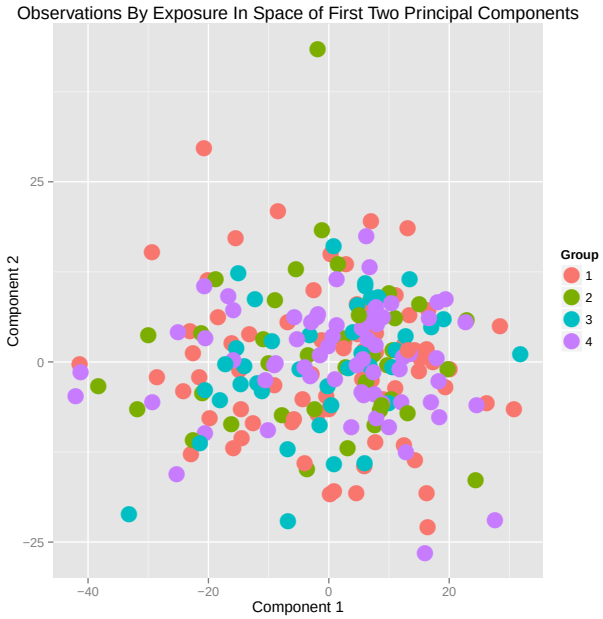


Figure 6: Disposition of Each Gene Chip[in the Space of the First Two Principal Components. Exposure Groups are Colour Coded.

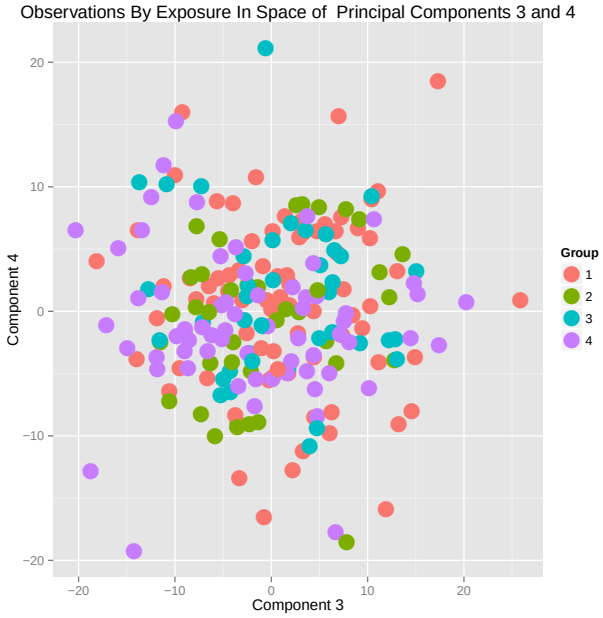


Figure 7: Disposition of Each Gene Chip in the Space of the Third and Fourth Principal Components. Exposure Groups are Colour Coded.

The gene expression data were clustered using three hierarchical clustering algorithms[2]: complete, average and single linkage[4]. Cluster plots are often used to over-interpret data. Apparent cluster structure is frequently illusory. Cluster solutions should only be interpreted when they are markedly consistent across different algorithms.

Figure 8 shows cluster solutions using all three algorithms. Single linkage produces chained clusters, in which each step in the hierarchical aggregation adds one more observation to an existing large cluster. Complete linkage produces nicely balanced solutions suggesting the presence of discrete groups. Unfortunately, this aesthetic preference for complete linkage does not reflect real structure in the data. The shape of the solution is markedly different between algorithms, strongly suggesting the absence of any true cluster structure.

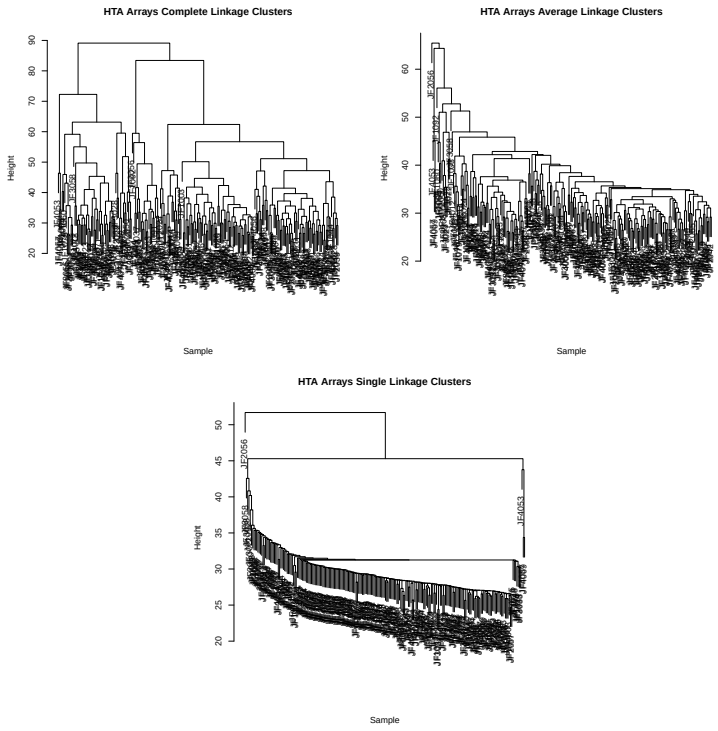


Figure 8: Cluster Analysis based on Complete Average and Single Linkage

The absence of any obvious structure in these exploratory analyses does not demonstrate that there are no differences between the groups. Rather it suggests that any differences are small relative to the inherent within group variation.

4 Differential Expression

4.1 Between Group Differences

Between group differences were evaluated with a linear model, and empirical Bayes moderated t tests[9] based on linear model parameter estimates. These were implemented using Smyth's limma methodology[13]. The following between group comparisons were performed:

1. Group 1 (Control) vs group 2 (Low exposure);
2. Group 1 (Control) vs group 3 (Moderate exposure);
3. Group 1 (Control) vs group 4 (High Exposure);

Date at which chips were scanned was included as a covariate (in an attempt to control for differences reflecting aging of reagents).

Individual p values for each gene were adjusted to maintain a false discovery rate of 5% using the Benjamini Hochberg procedure[15]. This procedure was adopted separately for each between groups comparison.

The probe sets were ranked in decreasing order of absolute $\log_2$ fold change, and are shown in Table 2.

Comparison	Entrez	Gene Symbol	Gene Name	logFC	AveExpr	t	adj.P.Val
High vs Control	151230	KLHL23	kelch-like family member 23	0.05	3.25	4.26	0.02
High vs Control	8812	CCNK	cyclin K	-0.06	8.86	-3.75	0.05
High vs Control	11055	ZPBP	zona pellucida binding protein	0.06	3.30	4.10	0.03
High vs Control	7428	VHL	von Hippel-Lindau tumor suppressor, E3 ubiquitin protein ligase	0.07	8.70	4.31	0.02
High vs Control	255324	EPGN	epithelial mitogen	0.07	2.98	3.75	0.05
High vs Control	5514	PPP1R10	protein phosphatase 1, regulatory subunit 10	-0.07	8.97	-3.96	0.03
High vs Control	5514	PPP1R10	protein phosphatase 1, regulatory subunit 10	-0.07	8.97	-3.96	0.03
High vs Control	5514	PPP1R10	protein phosphatase 1, regulatory subunit 10	-0.07	8.97	-3.96	0.03
High vs Control	5514	PPP1R10	protein phosphatase 1, regulatory subunit 10	-0.07	8.97	-3.96	0.03
High vs Control	5514	PPP1R10	protein phosphatase 1, regulatory subunit 10	-0.07	8.97	-3.96	0.03
High vs Control	55585	UBE2Q1	ubiquitin-conjugating enzyme E2Q family member 1	-0.07	8.94	-4.49	0.02
High vs Control	5514	PPP1R10	protein phosphatase 1, regulatory subunit 10	-0.07	8.96	-4.00	0.03
High vs Control	5514	PPP1R10	protein phosphatase 1, regulatory subunit 10	-0.07	8.86	-4.01	0.03
High vs Control	2733	GLE1	GLE1 RNA export mediator	-0.07	8.49	-4.49	0.02
High vs Control	54455	FBXO42	F-box protein 42	-0.08	8.12	-4.20	0.03
High vs Control	9898	UBAP2L	ubiquitin associated protein 2-like	-0.08	8.70	-3.74	0.05
High vs Control	29959	NRBP1	nuclear receptor binding protein 1	-0.08	8.90	-4.70	0.01
High vs Control	57168	ASPHD2	aspartate beta-hydroxylase domain containing 2	0.08	7.81	4.32	0.02
High vs Control	145853	C15orf61	chromosome 15 open reading frame 61	0.08	6.42	3.85	0.04
High vs Control	55146	ZDHHC4	zinc finger, DHHC-type containing 4	0.08	7.69	4.78	0.01
High vs Control	9352	TXNL1	thioredoxin-like 1	0.08	8.07	3.95	0.03
High vs Control	55170	PRMT6	protein arginine methyltransferase 6	0.09	6.94	3.90	0.04
High vs Control	100288730	PAN3-AS1	PAN3 antisense RNA 1	0.09	7.48	3.77	0.05

Table 2: Statistically Significant Differences in Gene Expression by Contrast; Results are ranked by absolute $\log_2$ fold change. Gene names and gene symbols are as defined by the HUGO Gene Nomenclature Committee `www.genenames.org`. Entrez is the NCBI Entrez code for the gene `http://www.ncbi.nlm.nih.gov/gene/`. Note some gene symbols are associated with multiple probe sets, and therefore have multiple entries in this table.

Comparison	Entrez	Gene Symbol	Gene Name	logFC	AveExpr	t	adj.P.Val
High vs Control	55252	ASXL2	additional sex combs like 2 (Drosophila)	-0.09	8.94	-4.16	0.03
High vs Control	2130	EWSR1	EWS RNA-binding protein 1	-0.09	8.37	-3.95	0.03
High vs Control	2145	EZH1	enhancer of zeste homolog 1 (Drosophila)	-0.09	9.49	-4.09	0.03
High vs Control	10482	NXF1	nuclear RNA export factor 1	-0.09	9.26	-4.11	0.03
High vs Control	161145	TMEM229B	transmembrane protein 229B	0.09	10.74	3.80	0.04
High vs Control	83593	RASSF5	Ras association (RalGDS/AF-6) domain family member 5	-0.09	10.23	-3.90	0.04
High vs Control	8841	HDAC3	histone deacetylase 3	-0.09	8.00	-3.86	0.04
High vs Control	50628	GEMIN4	gem (nuclear organelle) associated protein 4	0.09	6.80	3.90	0.04
High vs Control	5546	PRCC	papillary renal cell carcinoma (translocation-associated)	-0.09	7.80	-4.89	0.01
High vs Control	100873848	RN7SKP4	RNA, 7SK small nuclear pseudogene 4	0.09	6.85	3.86	0.04
High vs Control	54856	GON4L	gon-4-like (C. elegans)	-0.09	8.03	-3.76	0.05
High vs Control	100873850	RN7SKP6	RNA, 7SK small nuclear pseudogene 6	0.09	5.07	4.81	0.01
High vs Control	29089	UBE2T	ubiquitin-conjugating enzyme E2T (putative)	0.09	6.28	3.90	0.04
High vs Control	4802	NFYC	nuclear transcription factor Y, gamma	-0.09	8.34	-4.50	0.02
High vs Control	9798	IST1	increased sodium tolerance 1 homolog (yeast)	-0.09	9.47	-4.30	0.02
High vs Control	100873847	RN7SKP3	RNA, 7SK small nuclear pseudogene 3	0.10	9.20	3.92	0.04
High vs Control	343505	NBPF7	neuroblastoma breakpoint family, member 7	-0.10	5.38	-3.98	0.03
High vs Control	11016	ATF7	activating transcription factor 7	-0.10	7.90	-4.23	0.02
High vs Control	693140	MIR555	microRNA 555	-0.10	9.22	-3.95	0.03
High vs Control	6601	SMARCC2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily c, member 2	-0.10	8.86	-4.18	0.03
High vs Control	126074	SWSAP1	SWIM-type zinc finger 7 associated protein 1	0.10	6.70	3.90	0.04
High vs Control	100873873	RN7SKP10	RNA, 7SK small nuclear pseudogene 10	0.10	6.64	4.13	0.03

Table 2: Statistically Significant Differences in Gene Expression by Contrast; Results are ranked by absolute $\log_2$ fold change. Gene names and gene symbols are as defined by the HUGO Gene Nomenclature Committee `www.genenames.org`. Entrez is the NCBI Entrez code for the gene `http://www.ncbi.nlm.nih.gov/gene/`. Note some gene symbols are associated with multiple probe sets, and therefore have multiple entries in this table.

Comparison	Entrez	Gene Symbol	Gene Name	logFC	AveExpr	t	adj.P.Val
High vs Control	9815	GIT2	G protein-coupled receptor kinase interacting ArfGAP 2	-0.10	10.39	-3.75	0.05
High vs Control	29107	NXT1	NTF2-like export factor 1	0.10	6.85	4.32	0.02
High vs Control	146434	ZNF597	zinc finger protein 597	0.10	8.20	3.89	0.04
High vs Control	23471	TRAM1	translocation associated membrane protein 1	0.10	9.32	4.08	0.03
High vs Control	8021	NUP214	nucleoporin 214kDa	-0.10	8.91	-4.21	0.03
High vs Control	100423018	MIR3156-3	microRNA 3156-3	0.10	4.50	3.88	0.04
High vs Control	100616263	MIR4485	microRNA 4485	0.10	10.73	4.11	0.03
High vs Control	51335	NGRN	neugrin, neurite outgrowth associated	0.10	7.41	4.09	0.03
High vs Control	30011	SH3KBP1	SH3-domain kinase binding protein 1	-0.10	9.03	-4.08	0.03
High vs Control	286827	TRIM59	tripartite motif containing 59	0.10	6.90	3.74	0.05
High vs Control	100294145	LOC100294145	uncharacterized LOC100294145	0.10	7.27	3.91	0.04
High vs Control	9377	COX5A	cytochrome c oxidase subunit Va	0.11	8.07	3.74	0.05
High vs Control	8655	DYNLL1	dynein, light chain, LC8-type 1	0.11	7.42	4.58	0.02
High vs Control	100294145	LOC100294145	uncharacterized LOC100294145	0.11	6.65	3.85	0.04
High vs Control	94107	TMEM203	transmembrane protein 203	0.11	8.10	3.74	0.05
High vs Control	641373	NRON	non-protein coding RNA, repressor of NFAT	0.11	4.71	3.77	0.05
High vs Control	23339	VPS39	vacuolar protein sorting 39 homolog (S. cerevisiae)	-0.11	8.93	-3.85	0.04
High vs Control	9682	KDM4A	lysine (K)-specific demethylase 4A	-0.11	7.67	-4.00	0.03
High vs Control	51227	PIGP	phosphatidylinositol glycan anchor biosynthesis, class P	0.11	6.99	4.39	0.02
High vs Control	1659	DHX8	DEAH (Asp-Glu-Ala-His) box polypeptide 8	-0.11	8.44	-3.89	0.04
High vs Control	2720	GLB1	galactosidase, beta 1	-0.12	8.30	-4.33	0.02
High vs Control	79595	SAP130	Sin3A-associated protein, 130kDa	-0.12	8.92	-4.91	0.01
High vs Control	4717	NDUFC1	NADH dehydrogenase (ubiquinone) 1, subcomplex unknown, 1, 6kDa	0.12	6.79	5.69	0.00

Table 2: Statistically Significant Differences in Gene Expression by Contrast; Results are ranked by absolute $\log_2$ fold change. Gene names and gene symbols are as defined by the HUGO Gene Nomenclature Committee `www.genenames.org`. Entrez is the NCBI Entrez code for the gene `http://www.ncbi.nlm.nih.gov/gene/`. Note some gene symbols are associated with multiple probe sets, and therefore have multiple entries in this table.

Comparison	Entrez	Gene Symbol	Gene Name	logFC	AveExpr	t	adj.P.Val
High vs Control	51317	PHF21A	PHD finger protein 21A	-0.12	9.72	-3.75	0.05
High vs Control	100128055	SMARCA5-AS1	SMARCA5 antisense RNA 1	0.12	6.78	3.77	0.05
High vs Control	23389	MED13L	mediator complex subunit 13-like	-0.12	10.69	-4.09	0.03
High vs Control	23352	UBR4	ubiquitin protein ligase E3 component n-recognin 4	-0.12	8.91	-3.93	0.04
High vs Control	7298	TYMS	thymidylate synthetase	0.12	6.16	4.10	0.03
High vs Control	63916	ELMO2	engulfment and cell motility 2	-0.12	8.07	-4.44	0.02
High vs Control	100873851	RN7SKP7	RNA, 7SK small nuclear pseudogene 7	0.13	6.67	4.79	0.01
High vs Control	8729	GBF1	golgi brefeldin A resistant guanine nucleotide exchange factor 1	-0.13	8.36	-4.58	0.02
High vs Control	29937	NENF	neudesin neurotrophic factor	0.13	8.22	3.74	0.05
High vs Control	23126	POGZ	pogo transposable element with ZNF domain	-0.13	8.54	-4.45	0.02
High vs Control	1794	DOCK2	dedicator of cytokinesis 2	-0.13	10.23	-4.04	0.03
High vs Control	100132101	HERC2P7	hect domain and RLD 2 pseudogene 7	-0.13	7.12	-3.83	0.04
High vs Control	2197	FAU	Finkel-Biskis-Reilly murine sarcoma virus (FBR-MuSV) ubiquitously expressed	0.13	7.74	4.32	0.02
High vs Control	100132406	NBPF10	neuroblastoma breakpoint family, member 10	-0.13	11.80	-3.88	0.04
High vs Control	1431	CS	citrate synthase	-0.13	9.25	-3.91	0.04
High vs Control	23385	NCSTN	nicastrin	-0.13	8.55	-3.93	0.04
High vs Control	25840	METTL7A	methyltransferase like 7A	0.13	8.25	4.22	0.03
High vs Control	81704	DOCK8	dedicator of cytokinesis 8	-0.13	10.59	-3.95	0.03
High vs Control	3716	JAK1	Janus kinase 1	-0.14	10.67	-4.47	0.02
High vs Control	9611	NCOR1	nuclear receptor corepressor 1	-0.14	9.24	-4.04	0.03
High vs Control	55353	LAPTM4B	lysosomal protein transmembrane 4 beta	0.14	6.70	3.91	0.04
High vs Control	100132406	NBPF10	neuroblastoma breakpoint family, member 10	-0.14	11.60	-4.03	0.03
High vs Control	9844	ELMO1	engulfment and cell motility 1	-0.14	8.86	-4.58	0.02
High vs Control	25832	NBPF14	neuroblastoma breakpoint family, member 14	-0.14	11.44	-3.93	0.04

Table 2: Statistically Significant Differences in Gene Expression by Contrast; Results are ranked by absolute $\log_2$ fold change. Gene names and gene symbols are as defined by the HUGO Gene Nomenclature Committee `www.genenames.org`. Entrez is the NCBI Entrez code for the gene `http://www.ncbi.nlm.nih.gov/gene/`. Note some gene symbols are associated with multiple probe sets, and therefore have multiple entries in this table.

Comparison	Entrez	Gene Symbol	Gene Name	logFC	AveExpr	t	adj.P.Val
High vs Control	55672	NBPF1	neuroblastoma breakpoint family, member 1	-0.14	8.95	-3.73	0.05
High vs Control	100302195	MIR320C2	microRNA 320c-2	0.15	3.66	4.51	0.02
High vs Control	55187	VPS13D	vacuolar protein sorting 13 homolog D (S. cerevisiae)	-0.15	8.19	-3.79	0.05
High vs Control	2730	GCLM	glutamate-cysteine ligase, modifier subunit	0.15	7.01	3.98	0.03
High vs Control	100873871	RNU5A-6P	RNA, U5A small nuclear 6, pseudogene	0.15	3.06	4.27	0.02
High vs Control	645251	CBX1P1	chromobox homolog 1 pseudogene 1	-0.15	7.70	-3.99	0.03
High vs Control	574452	MIR494	microRNA 494	0.16	5.70	4.11	0.03
High vs Control	116540	MRPL53	mitochondrial ribosomal protein L53	0.16	8.12	4.53	0.02
High vs Control	400818	NBPF9	neuroblastoma breakpoint family, member 9	-0.17	10.57	-3.98	0.03
High vs Control	388677	NOTCH2NL	notch 2 N-terminal like	-0.17	10.66	-4.37	0.02
High vs Control	100129034	LOC100129034	uncharacterized LOC100129034	-0.17	9.50	-3.81	0.04
High vs Control	400818	NBPF9	neuroblastoma breakpoint family, member 9	-0.17	9.97	-4.07	0.03
High vs Control	401105	FLJ42393	uncharacterized LOC401105	-0.20	6.56	-4.07	0.03
High vs Control	6628	SNRPB	small nuclear ribonucleoprotein polypeptides B and B1	-0.20	7.97	-6.04	0.00
High vs Control	693208	MIR623	microRNA 623	-0.20	7.32	-4.14	0.03
High vs Control	8366	HIST1H4B	histone cluster 1, H4b	0.21	5.64	4.32	0.02
High vs Control	677819	SNORA37	small nucleolar RNA, H/ACA box 37	0.22	9.81	3.82	0.04
High vs Control	100616151	MIR4480	microRNA 4480	-0.23	6.87	-3.95	0.03
High vs Control	692225	SNORD94	small nucleolar RNA, C/D box 94	0.23	10.05	3.73	0.05
High vs Control	100151687	RNU6ATAC4P	RNA, U6atac small nuclear 4, pseudogene	0.28	5.02	3.97	0.03
High vs Control	100132288	TEKT4P2	tektin 4 pseudogene 2	0.29	8.30	3.81	0.04
High vs Control	693197	MIR612	microRNA 612	-0.30	10.27	-4.53	0.02
High vs Control	100151688	RNU6ATAC5P	RNA, U6atac small nuclear 5, pseudogene	0.40	6.11	3.79	0.05
High vs Control	100151686	RNU6ATAC3P	RNA, U6atac small nuclear 3, pseudogene	0.44	5.21	3.95	0.03
High vs Control	100151685	RNU6ATAC2P	RNA, U6atac small nuclear 2, pseudogene	0.44	7.52	4.50	0.02

Table 2: Statistically Significant Differences in Gene Expression by Contrast; Results are ranked by absolute $\log_2$ fold change. Gene names and gene symbols are as defined by the HUGO Gene Nomenclature Committee `www.genenames.org`. Entrez is the NCBI Entrez code for the gene `http://www.ncbi.nlm.nih.gov/gene/`. Note some gene symbols are associated with multiple probe sets, and therefore have multiple entries in this table.

Comparison	Entrez	Gene Symbol	Gene Name	logFC	AveExpr	t	adj.P.Val
High vs Control	100151684	RNU6ATAC	RNA, U6atac small nuclear (U12-dependent splicing)	0.58	8.81	4.19	0.03
Low vs Control	51072	MEMO1	mediator of cell motility 1	0.05	8.29	3.89	0.04
Low vs Control	142686	ASB14	ankyrin repeat and SOCS box containing 14	0.06	3.61	3.93	0.03
Low vs Control	151230	KLHL23	kelch-like family member 23	0.06	3.25	4.50	0.02
Low vs Control	126969	SLC44A3	solute carrier family 44, member 3	0.08	4.47	4.18	0.02
Low vs Control	5514	PPP1R10	protein phosphatase 1, regulatory subunit 10	-0.08	8.97	-3.72	0.05
Low vs Control	5514	PPP1R10	protein phosphatase 1, regulatory subunit 10	-0.08	8.97	-3.72	0.05
Low vs Control	5514	PPP1R10	protein phosphatase 1, regulatory subunit 10	-0.08	8.97	-3.72	0.05
Low vs Control	5514	PPP1R10	protein phosphatase 1, regulatory subunit 10	-0.08	8.97	-3.72	0.05
Low vs Control	5514	PPP1R10	protein phosphatase 1, regulatory subunit 10	-0.08	8.97	-3.72	0.05
Low vs Control	5514	PPP1R10	protein phosphatase 1, regulatory subunit 10	-0.08	8.96	-3.71	0.05
Low vs Control	100129385	C9orf92	chromosome 9 open reading frame 92	0.08	4.89	3.93	0.03
Low vs Control	64768	IPPK	inositol 1,3,4,5,6-pentakisphosphate 2-kinase	-0.08	7.03	-4.01	0.03
Low vs Control	114134	SLC2A13	solute carrier family 2 (facilitated glucose transporter), member 13	0.08	6.44	3.81	0.04
Low vs Control	5514	PPP1R10	protein phosphatase 1, regulatory subunit 10	-0.08	8.86	-3.82	0.04
Low vs Control	79147	FKRP	fukutin related protein	0.08	7.12	3.70	0.05
Low vs Control	5546	PRCC	papillary renal cell carcinoma (translocation-associated)	-0.09	7.80	-3.93	0.03
Low vs Control	124923	SGK494	uncharacterized serine/threonine-protein kinase SgK494	-0.09	6.44	-4.19	0.02
Low vs Control	100874215	ATP13A4-AS1	ATP13A4 antisense RNA 1	0.09	3.44	3.86	0.04
Low vs Control	9203	ZMYM3	zinc finger, MYM-type 3	-0.09	7.02	-3.78	0.05
Low vs Control	22985	ACIN1	apoptotic chromatin condensation inducer 1	-0.09	8.75	-4.03	0.03
Low vs Control	219902	TMEM136	transmembrane protein 136	0.09	4.63	3.78	0.05
Low vs Control	55146	ZDHHC4	zinc finger, DHHC-type containing 4	0.10	7.69	4.70	0.02

Table 2: Statistically Significant Differences in Gene Expression by Contrast; Results are ranked by absolute $\log_2$ fold change. Gene names and gene symbols are as defined by the HUGO Gene Nomenclature Committee `www.genenames.org`. Entrez is the NCBI Entrez code for the gene `http://www.ncbi.nlm.nih.gov/gene/`. Note some gene symbols are associated with multiple probe sets, and therefore have multiple entries in this table.

Comparison	Entrez	Gene Symbol	Gene Name	logFC	AveExpr	t	adj.P.Val
Low vs Control	2145	EZH1	enhancer of zeste homolog 1 (Drosophila)	-0.10	9.49	-3.78	0.05
Low vs Control	163126	EID2	EP300 interacting inhibitor of differentiation 2	0.10	7.94	3.71	0.05
Low vs Control	84260	TCHP	trichoplein, keratin filament binding	-0.10	7.45	-4.12	0.02
Low vs Control	8655	DYNLL1	dynein, light chain, LC8-type 1	0.10	7.42	3.74	0.05
Low vs Control	55252	ASXL2	additional sex combs like 2 (Drosophila)	-0.10	8.94	-4.11	0.02
Low vs Control	161145	TMEM229B	transmembrane protein 229B	0.10	10.74	3.72	0.05
Low vs Control	9681	DEPDC5	DEP domain containing 5	-0.11	7.87	-3.79	0.05
Low vs Control	50628	GEMIN4	gem (nuclear organelle) associated protein 4	0.11	6.80	3.75	0.05
Low vs Control	223082	ZNRF2	zinc and ring finger 2	0.11	8.63	4.50	0.02
Low vs Control	29107	NXT1	NTF2-like export factor 1	0.11	6.85	3.88	0.04
Low vs Control	4717	NDUFC1	NADH dehydrogenase (ubiquinone) 1, subcomplex unknown, 1, 6kDa	0.11	6.79	4.41	0.02
Low vs Control	6612	SUMO3	small ubiquitin-like modifier 3	0.11	8.25	4.18	0.02
Low vs Control	26248	OR2K2	olfactory receptor, family 2, subfamily K, member 2	0.11	2.71	3.81	0.04
Low vs Control	84539	MCHR2	melanin-concentrating hormone receptor 2	0.11	4.51	3.91	0.04
Low vs Control	128367	OR10X1	olfactory receptor, family 10, subfamily X, member 1	0.11	2.53	4.24	0.02
Low vs Control	408187	SPINK14	serine peptidase inhibitor, Kazal type 14 (putative)	0.11	3.04	3.73	0.05
Low vs Control	693140	MIR555	microRNA 555	-0.11	9.22	-3.79	0.05
Low vs Control	6293	VPS52	vacuolar protein sorting 52 homolog (S. cerevisiae)	-0.11	7.55	-3.70	0.05
Low vs Control	6293	VPS52	vacuolar protein sorting 52 homolog (S. cerevisiae)	-0.11	7.55	-3.70	0.05
Low vs Control	647042	GOLGA6L10	golgin A6 family-like 10	-0.11	6.74	-3.94	0.03

Table 2: Statistically Significant Differences in Gene Expression by Contrast; Results are ranked by absolute $\log_2$ fold change. Gene names and gene symbols are as defined by the HUGO Gene Nomenclature Committee `www.genenames.org`. Entrez is the NCBI Entrez code for the gene `http://www.ncbi.nlm.nih.gov/gene/`. Note some gene symbols are associated with multiple probe sets, and therefore have multiple entries in this table.

Comparison	Entrez	Gene Symbol	Gene Name	logFC	AveExpr	t	adj.P.Val
Low vs Control	100506710	LOC100506710	endogenous Bornavirus-like nucleoprotein 2 pseudogene	0.11	8.00	4.36	0.02
Low vs Control	6293	VPS52	vacuolar protein sorting 52 homolog (S. cerevisiae)	-0.11	7.55	-3.76	0.05
Low vs Control	6293	VPS52	vacuolar protein sorting 52 homolog (S. cerevisiae)	-0.11	7.55	-3.76	0.05
Low vs Control	6293	VPS52	vacuolar protein sorting 52 homolog (S. cerevisiae)	-0.11	7.55	-3.77	0.05
Low vs Control	56850	GRIPAP1	GRIP1 associated protein 1	-0.11	7.64	-4.00	0.03
Low vs Control	100873845	RN7SKP1	RNA, 7SK small nuclear pseudogene 1	0.11	6.76	4.05	0.03
Low vs Control	51504	TRMT112	tRNA methyltransferase 11-2 homolog (S. cerevisiae)	0.12	9.31	3.76	0.05
Low vs Control	8021	NUP214	nucleoporin 214kDa	-0.12	8.91	-4.11	0.02
Low vs Control	11016	ATF7	activating transcription factor 7	-0.12	7.90	-4.37	0.02
Low vs Control	3660	IRF2	interferon regulatory factor 2	-0.12	9.80	-4.04	0.03
Low vs Control	146434	ZNF597	zinc finger protein 597	0.12	8.20	3.92	0.04
Low vs Control	79595	SAP130	Sin3A-associated protein, 130kDa	-0.12	8.92	-4.34	0.02
Low vs Control	23112	TNRC6B	trinucleotide repeat containing 6B	-0.12	9.26	-4.03	0.03
Low vs Control	92483	LDHAL6B	lactate dehydrogenase A-like 6B	0.12	2.78	4.63	0.02
Low vs Control	6601	SMARCC2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily c, member 2	-0.12	8.86	-4.45	0.02
Low vs Control	2130	EWSR1	EWS RNA-binding protein 1	-0.12	8.37	-4.62	0.02
Low vs Control	100380271	RNY5P4	RNA, Ro-associated Y5 pseudogene 4	0.12	4.22	3.97	0.03
Low vs Control	51335	NGRN	neugrin, neurite outgrowth associated	0.12	7.41	4.15	0.02
Low vs Control	116138	KLHDC3	kelch domain containing 3	0.13	8.48	4.32	0.02
Low vs Control	343505	NBPF7	neuroblastoma breakpoint family, member 7	-0.13	5.38	-4.40	0.02

Table 2: Statistically Significant Differences in Gene Expression by Contrast; Results are ranked by absolute $\log_2$ fold change. Gene names and gene symbols are as defined by the HUGO Gene Nomenclature Committee `www.genenames.org`. Entrez is the NCBI Entrez code for the gene `http://www.ncbi.nlm.nih.gov/gene/`. Note some gene symbols are associated with multiple probe sets, and therefore have multiple entries in this table.

Comparison	Entrez	Gene Symbol	Gene Name	logFC	AveExpr	t	adj.P.Val
Low vs Control	440295	GOLGA6L9	golgin A6 family-like 9	-0.13	6.64	-4.15	0.02
Low vs Control	131616	TMEM42	transmembrane protein 42	0.13	6.70	3.78	0.05
Low vs Control	100294145	LOC100294145	uncharacterized LOC100294145	0.13	7.27	4.01	0.03
Low vs Control	51227	PIGP	phosphatidylinositol glycan anchor biosynthesis, class P	0.13	6.99	4.34	0.02
Low vs Control	9240	PNMA1	paraneoplastic Ma antigen 1	0.13	6.96	3.76	0.05
Low vs Control	23471	TRAM1	translocation associated membrane protein 1	0.13	9.32	4.46	0.02
Low vs Control	55266	TMEM19	transmembrane protein 19	0.13	7.85	3.75	0.05
Low vs Control	54856	GON4L	gon-4-like (C. elegans)	-0.13	8.03	-4.51	0.02
Low vs Control	440295	GOLGA6L9	golgin A6 family-like 9	-0.13	6.29	-4.05	0.03
Low vs Control	2800	GOLGA1	golgin A1	-0.14	7.34	-4.80	0.02
Low vs Control	94107	TMEM203	transmembrane protein 203	0.14	8.10	3.95	0.03
Low vs Control	9682	KDM4A	lysine (K)-specific demethylase 4A	-0.14	7.67	-4.24	0.02
Low vs Control	54407	SLC38A2	solute carrier family 38, member 2	0.14	10.57	4.47	0.02
Low vs Control	100302237	MIR1281	microRNA 1281	-0.14	9.79	-4.03	0.03
Low vs Control	23126	POGZ	pogo transposable element with ZNF domain	-0.14	8.54	-4.06	0.03
Low vs Control	23339	VPS39	vacuolar protein sorting 39 homolog (S. cerevisiae)	-0.14	8.93	-4.15	0.02
Low vs Control	1659	DHX8	DEAH (Asp-Glu-Ala-His) box polypeptide 8	-0.14	8.44	-4.03	0.03
Low vs Control	11176	BAZ2A	bromodomain adjacent to zinc finger domain, 2A	-0.14	9.22	-3.84	0.04
Low vs Control	9869	SETDB1	SET domain, bifurcated 1	-0.14	7.65	-4.00	0.03
Low vs Control	23215	PRRC2C	proline-rich coiled-coil 2C	-0.14	10.22	-3.80	0.05
Low vs Control	9844	ELMO1	engulfment and cell motility 1	-0.15	8.86	-3.98	0.03
Low vs Control	23352	UBR4	ubiquitin protein ligase E3 component n-recognin 4	-0.15	8.91	-4.09	0.03
Low vs Control	100294145	LOC100294145	uncharacterized LOC100294145	0.15	7.21	3.90	0.04

Table 2: Statistically Significant Differences in Gene Expression by Contrast; Results are ranked by absolute $\log_2$ fold change. Gene names and gene symbols are as defined by the HUGO Gene Nomenclature Committee `www.genenames.org`. Entrez is the NCBI Entrez code for the gene `http://www.ncbi.nlm.nih.gov/gene/`. Note some gene symbols are associated with multiple probe sets, and therefore have multiple entries in this table.

Comparison	Entrez	Gene Symbol	Gene Name	logFC	AveExpr	t	adj.P.Val
Low vs Control	100294145	LOC100294145	uncharacterized LOC100294145	0.15	7.21	3.90	0.04
Low vs Control	100294145	LOC100294145	uncharacterized LOC100294145	0.15	7.21	3.90	0.04
Low vs Control	4259	MGST3	microsomal glutathione S-transferase 3	0.15	6.46	3.90	0.04
Low vs Control	81704	DOCK8	dedicator of cytokinesis 8	-0.15	10.59	-3.76	0.05
Low vs Control	100294145	LOC100294145	uncharacterized LOC100294145	0.15	6.65	4.63	0.02
Low vs Control	8729	GBF1	golgi brefeldin A resistant guanine nucleotide exchange factor 1	-0.16	8.36	-4.74	0.02
Low vs Control	55421	C17orf85	chromosome 17 open reading frame 85	-0.16	8.39	-3.90	0.04
Low vs Control	1794	DOCK2	dedicator of cytokinesis 2	-0.16	10.23	-4.15	0.02
Low vs Control	100294145	LOC100294145	uncharacterized LOC100294145	0.16	7.19	4.47	0.02
Low vs Control	100294145	LOC100294145	uncharacterized LOC100294145	0.16	6.86	4.43	0.02
Low vs Control	100294145	LOC100294145	uncharacterized LOC100294145	0.16	6.86	4.43	0.02
Low vs Control	23420	NOMO1	NODAL modulator 1	-0.17	9.69	-4.18	0.02
Low vs Control	9611	NCOR1	nuclear receptor corepressor 1	-0.17	9.24	-4.15	0.02
Low vs Control	8243	SMC1A	structural maintenance of chromosomes 1A	-0.17	8.26	-3.83	0.04
Low vs Control	55672	NBPF1	neuroblastoma breakpoint family, member 1	-0.17	8.95	-3.71	0.05
Low vs Control	100132406	NBPF10	neuroblastoma breakpoint family, member 10	-0.17	11.80	-4.23	0.02
Low vs Control	100616498	MIR378E	microRNA 378e	-0.17	9.15	-4.21	0.02
Low vs Control	283820	NOMO2	NODAL modulator 2	-0.17	9.68	-4.22	0.02
Low vs Control	408050	NOMO3	NODAL modulator 3	-0.18	9.69	-4.25	0.02
Low vs Control	6830	SUPT6H	suppressor of Ty 6 homolog (S. cerevisiae)	-0.18	8.82	-4.37	0.02
Low vs Control	25832	NBPF14	neuroblastoma breakpoint family, member 14	-0.18	11.44	-4.07	0.03
Low vs Control	147226	UPF3AP1	UPF3A pseudogene 1	-0.18	7.20	-4.28	0.02
Low vs Control	6628	SNRPB	small nuclear ribonucleoprotein polypeptides B and B1	-0.18	7.97	-4.46	0.02
Low vs Control	100873800	RNY4P22	RNA, Ro-associated Y4 pseudogene 22	0.18	4.60	3.85	0.04

Table 2: Statistically Significant Differences in Gene Expression by Contrast; Results are ranked by absolute $\log_2$ fold change. Gene names and gene symbols are as defined by the HUGO Gene Nomenclature Committee `www.genenames.org`. Entrez is the NCBI Entrez code for the gene `http://www.ncbi.nlm.nih.gov/gene/`. Note some gene symbols are associated with multiple probe sets, and therefore have multiple entries in this table.

Comparison	Entrez	Gene Symbol	Gene Name	logFC	AveExpr	t	adj.P.Val
Low vs Control	26497	OR10D3	olfactory receptor, family 10, subfamily D, member 3 (non-functional)	0.18	6.35	4.29	0.02
Low vs Control	100132406	NBPF10	neuroblastoma breakpoint family, member 10	-0.18	11.60	-4.35	0.02
Low vs Control	652966	SNORD10	small nucleolar RNA, C/D box 10	0.19	9.07	3.72	0.05
Low vs Control	388677	NOTCH2NL	notch 2 N-terminal like	-0.19	10.66	-4.14	0.02
Low vs Control	400818	NBPF9	neuroblastoma breakpoint family, member 9	-0.19	10.57	-3.89	0.04
Low vs Control	100616319	MIR4493	microRNA 4493	0.20	4.30	4.38	0.02
Low vs Control	400818	NBPF9	neuroblastoma breakpoint family, member 9	-0.20	9.97	-3.92	0.04
Low vs Control	319103	SNORD8	small nucleolar RNA, C/D box 8	0.21	7.07	4.03	0.03
Low vs Control	767565	SNORD113-5	small nucleolar RNA, C/D box 113-5	0.22	3.98	3.98	0.03
Low vs Control	401105	FLJ42393	uncharacterized LOC401105	-0.22	6.56	-3.77	0.05
Low vs Control	693189	MIR604	microRNA 604	-0.22	9.46	-3.75	0.05
Low vs Control	131583	FAM43A	family with sequence similarity 43, member A	0.23	7.92	5.02	0.02
Low vs Control	677842	SNORA76	small nucleolar RNA, H/ACA box 76	0.25	6.87	3.91	0.04
Low vs Control	693208	MIR623	microRNA 623	-0.26	7.32	-4.32	0.02
Low vs Control	692225	SNORD94	small nucleolar RNA, C/D box 94	0.29	10.05	3.83	0.04
Low vs Control	100132686	LOC100132686	uncharacterized LOC100132686	-0.32	5.03	-4.76	0.02
Low vs Control	100616151	MIR4480	microRNA 4480	-0.33	6.87	-4.71	0.02
Low vs Control	693197	MIR612	microRNA 612	-0.36	10.27	-4.53	0.02
Low vs Control	100151687	RNU6ATAC4P	RNA, U6atac small nuclear 4, pseudogene	0.37	5.02	4.29	0.02
Low vs Control	100151688	RNU6ATAC5P	RNA, U6atac small nuclear 5, pseudogene	0.49	6.11	3.82	0.04
Low vs Control	100151685	RNU6ATAC2P	RNA, U6atac small nuclear 2, pseudogene	0.50	7.52	4.25	0.02
Low vs Control	100151684	RNU6ATAC	RNA, U6atac small nuclear (U12-dependent splicing)	0.68	8.81	4.11	0.02
Medium vs Control	84224	NBPF3	neuroblastoma breakpoint family, member 3	-0.14	7.68	-4.63	0.05
Medium vs Control	388677	NOTCH2NL	notch 2 N-terminal like	-0.21	10.66	-4.74	0.05

Table 2: Statistically Significant Differences in Gene Expression by Contrast; Results are ranked by absolute $\log_2$ fold change. Gene names and gene symbols are as defined by the HUGO Gene Nomenclature Committee `www.genenames.org`. Entrez is the NCBI Entrez code for the gene `http://www.ncbi.nlm.nih.gov/gene/`. Note some gene symbols are associated with multiple probe sets, and therefore have multiple entries in this table.

Comparison	Entrez	Gene Symbol	Gene Name	logFC	AveExpr	t	adj.P.Val
Medium vs Control	400818	NBPF9	neuroblastoma breakpoint family, member 9	-0.23	9.97	-4.58	0.05
Medium vs Control	100151687	RNU6ATAC4P	RNA, U6atac small nuclear 4, pseudogene	0.42	5.02	4.92	0.04

Table 2: Statistically Significant Differences in Gene Expression by Contrast; Results are ranked by absolute $\log_2$ fold change. Gene names and gene symbols are as defined by the HUGO Gene Nomenclature Committee `www.genenames.org`. Entrez is the NCBI Entrez code for the gene `http://www.ncbi.nlm.nih.gov/gene/`. Note some gene symbols are associated with multiple probe sets, and therefore have multiple entries in this table.

We note that despite the statistically significant results, the log fold changes are very small. The largest effect represents a 1.5 fold reduction in gene expression. many acute conditions will show gene expression changes of the order of 8 to 32 fold.

Statistically significant fold changes are compared across contrasts for each gene in Table 3. In general, the signs of changes are the same, though there is little evidence of a dose response. That is the difference between the high dose and control is not generally greater than the difference beween the low dose and the control.

Gene ID	Name	Low vs Control	Medium vs Control	High vs Control
SMARCA5-AS1	SMARCA5 antisense RNA 1	0.118		
LOC100129034	uncharacterized LOC100129034	-0.169		
C9orf92	chromosome 9 open reading frame 92			0.082
HERC2P7	hect domain and RLD 2 pseudogene 7	-0.130		
TEKT4P2	tektin 4 pseudogene 2	0.291		
NBPF10	neuroblastoma breakpoint family, member 10	-0.136		-0.177
LOC100132686	uncharacterized LOC100132686			-0.316
RNU6ATAC	RNA, U6atac small nuclear (U12-dependent splicing)	0.579		0.681
RNU6ATAC2P	RNA, U6atac small nuclear 2, pseudogene	0.438		0.495
RNU6ATAC3P	RNA, U6atac small nuclear 3, pseudogene	0.437		
RNU6ATAC4P	RNA, U6atac small nuclear 4, pseudogene	0.283	0.416	0.367
RNU6ATAC5P	RNA, U6atac small nuclear 5, pseudogene	0.404		0.489
PAN3-AS1	PAN3 antisense RNA 1	0.088		
LOC100294145	uncharacterized LOC100294145	0.106		0.152
MIR320C2	microRNA 320c-2	0.147		
MIR1281	microRNA 1281			-0.139
RNY5P4	RNA, Ro-associated Y5 pseudogene 4			0.124
MIR3156-3	microRNA 3156-3	0.101		
LOC100506710	endogenous Bornavirus-like nucleoprotein 2 pseudogene			0.112
MIR4480	microRNA 4480	-0.233		-0.334
MIR4485	microRNA 4485	0.101		
MIR4493	microRNA 4493			0.199
MIR378E	microRNA 378e			-0.173

Table 3: log2 Fold Change of mRNA Species By Comparison. Gene is the HUGO Gene Nomenclature Committee `www.genenames.org` Name for the gene.

Gene ID	Name	Low vs Control	Medium vs Control	High vs Control
RNY4P22	RNA, Ro-associated Y4 pseudogene 22			0.181
RN7SKP1	RNA, 7SK small nuclear pseudogene 1			0.114
RN7SKP3	RNA, 7SK small nuclear pseudogene 3	0.096		
RN7SKP4	RNA, 7SK small nuclear pseudogene 4	0.092		
RN7SKP6	RNA, 7SK small nuclear pseudogene 6	0.092		
RN7SKP7	RNA, 7SK small nuclear pseudogene 7	0.125		
RNU5A-6P	RNA, U5A small nuclear 6, pseudogene	0.152		
RN7SKP10	RNA, 7SK small nuclear pseudogene 10	0.098		
ATP13A4-AS1	ATP13A4 antisense RNA 1			0.091
NXF1	nuclear RNA export factor 1	-0.089		
ATF7	activating transcription factor 7	-0.096		-0.119
ZPBP	zona pellucida binding protein	0.059		
BAZ2A	bromodomain adjacent to zinc finger domain, 2A			-0.141
SLC2A13	solute carrier family 2 (facilitated glucose transporter), member 13			0.084
KLHDC3	kelch domain containing 3			0.125
MRPL53	mitochondrial ribosomal protein L53	0.165		
SGK494	uncharacterized serine/threonine-protein kinase SgK494			-0.089
SWSAP1	SWIM-type zinc finger 7 associated protein 1	0.097		
SLC44A3	solute carrier family 44, member 3			0.079
OR10X1	olfactory receptor, family 10, subfamily X, member 1			0.110
FAM43A	family with sequence similarity 43, member A			0.233

Table 3: log2 Fold Change of mRNA Species By Comparison. Gene is the HUGO Gene Nomenclature Committee `www.genenames.org` Name for the gene.

Gene ID	Name	Low vs Control	Medium vs Control	High vs Control
TMEM42	transmembrane protein 42			0.128
ASB14	ankyrin repeat and SOCS box containing 14			0.060
CS	citrate synthase	-0.133		
C15orf61	chromosome 15 open reading frame 61	0.079		
ZNF597	zinc finger protein 597	0.101		0.121
UPF3AP1	UPF3A pseudogene 1			-0.177
KLHL23	kelch-like family member 23	0.051		0.064
TMEM229B	transmembrane protein 229B	0.089		0.105
EID2	EP300 interacting inhibitor of differentiation 2			0.102
DHX8	DEAH (Asp-Glu-Ala-His) box polypeptide 8	-0.114		-0.141
DOCK2	dedicator of cytokinesis 2	-0.130		-0.160
EWSR1	EWS RNA-binding protein 1	-0.089		-0.124
EZH1	enhancer of zeste homolog 1 (Drosophila)	-0.089		-0.099
FAU	Finkel-Biskis-Reilly murine sarcoma virus (FBR-MuSV) ubiquitously expressed	0.131		
TMEM136	transmembrane protein 136			0.095
ZNRF2	zinc and ring finger 2			0.106
ACIN1	apoptotic chromatin condensation inducer 1			-0.093
TNRC6B	trinucleotide repeat containing 6B			-0.124
POGZ	pogo transposable element with ZNF domain	-0.128		-0.140
PRRC2C	proline-rich coiled-coil 2C			-0.143
VPS39	vacuolar protein sorting 39 homolog (S. cerevisiae)	-0.109		-0.141
UBR4	ubiquitin protein ligase E3 component n-recognin 4	-0.119		-0.149

Table 3: log2 Fold Change of mRNA Species By Comparison. Gene is the HUGO Gene Nomenclature Committee `www.genenames.org` Name for the gene.

Gene ID	Name	Low vs Control	Medium vs Control	High vs Control
NCSTN	nicastrin	-0.133		
MED13L	mediator complex subunit 13-like	-0.119		
NOMO1	NODAL modulator 1			-0.167
TRAM1	translocation associated membrane protein 1	0.101		0.132
EPGN	epithelial mitogen	0.070		
NBPF14	neuroblastoma breakpoint family, member 14	-0.142		-0.176
METTL7A	methyltransferase like 7A	0.134		
OR2K2	olfactory receptor, family 2, subfamily K, member 2			0.109
OR10D3	olfactory receptor, family 10, subfamily D, member 3 (non-functional)			0.182
GLB1	galactosidase, beta 1	-0.116		
GCLM	glutamate-cysteine ligase, modifier subunit	0.148		
GLE1	GLE1 RNA export mediator	-0.074		
GOLGA1	golgin A1			-0.136
NOMO2	NODAL modulator 2			-0.173
TRIM59	tripartite motif containing 59	0.104		
UBE2T	ubiquitin-conjugating enzyme E2T (putative)	0.093		
NXT1	NTF2-like export factor 1	0.100		0.107
NENF	neudesin neurotrophic factor	0.127		
NRBP1	nuclear receptor binding protein 1	-0.076		
SH3KBP1	SH3-domain kinase binding protein 1	-0.103		
SNORD8	small nucleolar RNA, C/D box 8			0.207
NBPF7	neuroblastoma breakpoint family, member 7	-0.096		-0.127
IRF2	interferon regulatory factor 2			-0.121
JAK1	Janus kinase 1	-0.136		

Table 3: log2 Fold Change of mRNA Species By Comparison. Gene is the HUGO Gene Nomenclature Committee `www.genenames.org` Name for the gene.

Gene ID	Name	Low vs Control	Medium vs Control	High vs Control
NOTCH2NL	notch 2 N-terminal like	-0.167	-0.214	-0.189
NBPF9	neuroblastoma breakpoint family, member 9	-0.170	-0.233	-0.198
FLJ42393	uncharacterized LOC401105	-0.195		-0.217
NOMO3	NODAL modulator 3			-0.176
SPINK14	serine peptidase inhibitor, Kazal type 14 (putative)			0.110
MGST3	microsomal glutathione S-transferase 3			0.153
GOLGA6L9	golgin A6 family-like 9			-0.130
NDUFC1	NADH dehydrogenase (ubiquinone) 1, subcomplex unknown, 1, 6kDa	0.117		0.108
NFYC	nuclear transcription factor Y, gamma	-0.095		
GEMIN4	gem (nuclear organelle) associated protein 4	0.092		0.105
MEMO1	mediator of cell motility 1			0.053
PIGP	phosphatidylinositol glycan anchor biosynthesis, class P	0.109		0.130
PHF21A	PHD finger protein 21A	-0.118		
NGRN	neugrin, neurite outgrowth associated	0.103		0.125
TRMT112	tRNA methyltransferase 11-2 homolog (S. cerevisiae)			0.118
SLC38A2	solute carrier family 38, member 2			0.139
FBXO42	F-box protein 42	-0.075		
GON4L	gon-4-like (C. elegans)	-0.092		-0.132
PPP1R10	protein phosphatase 1, regulatory subunit 10	-0.072		-0.081
ZDHHC4	zinc finger, DHHC-type containing 4	0.083		0.098
PRMT6	protein arginine methyltransferase 6	0.086		

Table 3: log2 Fold Change of mRNA Species By Comparison. Gene is the HUGO Gene Nomenclature Committee `www.genenames.org` Name for the gene.

Gene ID	Name	Low vs Control	Medium vs Control	High vs Control
VPS13D	vacuolar protein sorting 13 homolog D (S. cerevisiae)	-0.147		
ASXL2	additional sex combs like 2 (Drosophila)	-0.088		-0.105
TMEM19	transmembrane protein 19			0.132
LAPTM4B	lysosomal protein transmembrane 4 beta	0.138		
C17orf85	chromosome 17 open reading frame 85			-0.158
PRCC	papillary renal cell carcinoma (translocation-associated)	-0.092		-0.089
UBE2Q1	ubiquitin-conjugating enzyme E2Q family member 1	-0.072		
NBPF1	neuroblastoma breakpoint family, member 1	-0.143		-0.171
GRIPAP1	GRIP1 associated protein 1			-0.114
ASPHD2	aspartate beta-hydroxylase domain containing 2	0.078		
MIR494	microRNA 494	0.160		
VPS52	vacuolar protein sorting 52 homolog (S. cerevisiae)			-0.113
ELMO2	engulfment and cell motility 2	-0.125		
NRON	non-protein coding RNA, repressor of NFAT	0.109		
CBX1P1	chromobox homolog 1 pseudogene 1	-0.154		
GOLGA6L10	golgin A6 family-like 10			-0.112
IPPK	inositol 1,3,4,5,6-pentakisphosphate 2-kinase			-0.083
SNORD10	small nucleolar RNA, C/D box 10			0.189
SMARCC2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily c, member 2	-0.097		-0.124

Table 3: log2 Fold Change of mRNA Species By Comparison. Gene is the HUGO Gene Nomenclature Committee `www.genenames.org` Name for the gene.

Gene ID	Name	Low vs Control	Medium vs Control	High vs Control
SUMO3	small ubiquitin-like modifier 3			0.108
SNRPB	small nuclear ribonucleoprotein polypeptides B and B1	-0.202		-0.179
SNORA37	small nucleolar RNA, H/ACA box 37	0.223		
SNORA76	small nucleolar RNA, H/ACA box 76			0.246
SUPT6H	suppressor of Ty 6 homolog (S. cerevisiae)			-0.176
SNORD94	small nucleolar RNA, C/D box 94	0.234		0.288
MIR555	microRNA 555	-0.097		-0.111
MIR604	microRNA 604			-0.223
MIR612	microRNA 612	-0.300		-0.360
MIR623	microRNA 623	-0.205		-0.256
TYMS	thymidylate synthetase	0.122		
VHL	von Hippel-Lindau tumor suppressor, E3 ubiquitin protein ligase	0.069		
SNORD113-5	small nucleolar RNA, C/D box 113-5			0.216
FKRP	fukutin related protein			0.085
SAP130	Sin3A-associated protein, 130kDa	-0.116		-0.123
NUP214	nucleoporin 214kDa	-0.101		-0.118
DOCK8	dedicator of cytokinesis 8	-0.135		-0.154
SMC1A	structural maintenance of chromosomes 1A			-0.170
RASSF5	Ras association (RalGDS/AF-6) domain family member 5	-0.091		
HIST1H4B	histone cluster 1, H4b	0.211		
NBPF3	neuroblastoma breakpoint family, member 3		-0.143	
TCHP	trichoplein, keratin filament binding			-0.104
MCHR2	melanin-concentrating hormone receptor 2			0.110

Table 3: log2 Fold Change of mRNA Species By Comparison. Gene is the HUGO Gene Nomenclature Committee `www.genenames.org` Name for the gene.

Gene ID	Name	Low vs Control	Medium vs Control	High vs Control
DYNLL1	dynein, light chain, LC8-type 1	0.107		0.104
GBF1	golgi brefeldin A resistant guanine nucleotide exchange factor 1	-0.126		-0.156
CCNK	cyclin K	-0.055		
HDAC3	histone deacetylase 3	-0.091		
ZMYM3	zinc finger, MYM-type 3			-0.091
PNMA1	paraneoplastic Ma antigen 1			0.131
LDHAL6B	lactate dehydrogenase A-like 6B			0.124
TXNL1	thioredoxin-like 1	0.084		
COX5A	cytochrome c oxidase subunit Va	0.105		
TMEM203	transmembrane protein 203	0.108		0.136
NCOR1	nuclear receptor corepressor 1	-0.138		-0.170
DEPDC5	DEP domain containing 5			-0.105
KDM4A	lysine (K)-specific demethylase 4A	-0.109		-0.139
IST1	increased sodium tolerance 1 homolog (yeast)	-0.095		
GIT2	G protein-coupled receptor kinase interacting ArfGAP 2	-0.099		
ELMO1	engulfment and cell motility 1	-0.142		-0.147
SETDB1	SET domain, bifurcated 1			-0.143
UBAP2L	ubiquitin associated protein 2-like	-0.076		

Table 3: log2 Fold Change of mRNA Species By Comparison. Gene is the HUGO Gene Nomenclature Committee `www.genenames.org` Name for the gene.

Expression was plotted against exposure group for the genes associated with the 20 probe sets showing the largest fold changes for the comparison of the high exposure group with the controls. These plots are shown in Figures **??** to **??**. As expected, the differences are small, and largely confined to exposure group 4.

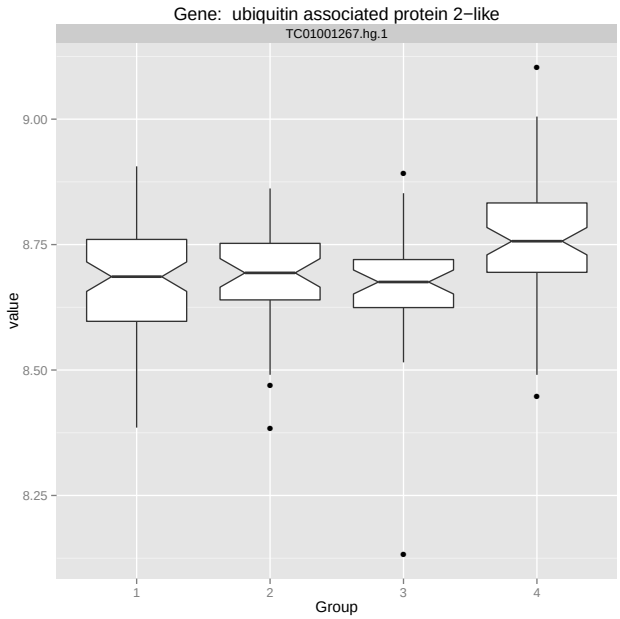


Figure 9: Group Differences for Gene Symbol UBAP2L : ubiquitin associated protein 2-like

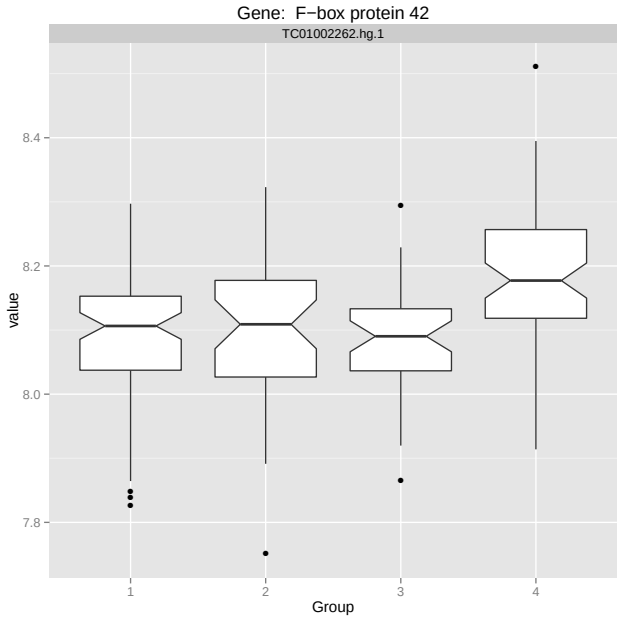


Figure 10: Group Differences for Gene Symbol FBXO42 : F-box protein 42

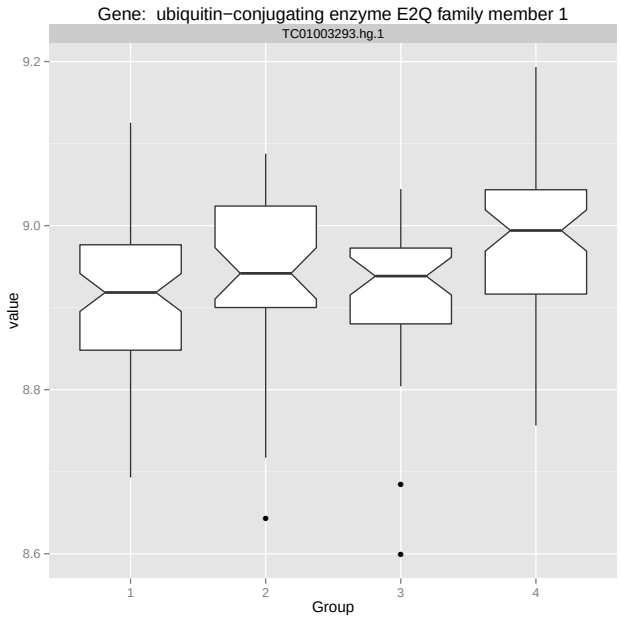


Figure 11: Group Differences for Gene Symbol UBE2Q1 : ubiquitin-conjugating enzyme E2Q family member 1

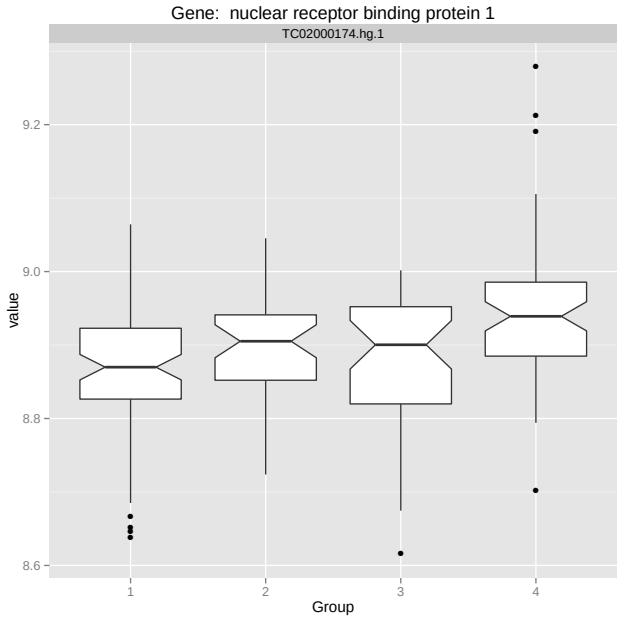


Figure 12: Group Differences for Gene Symbol NRBP1 : nuclear receptor binding protein 1

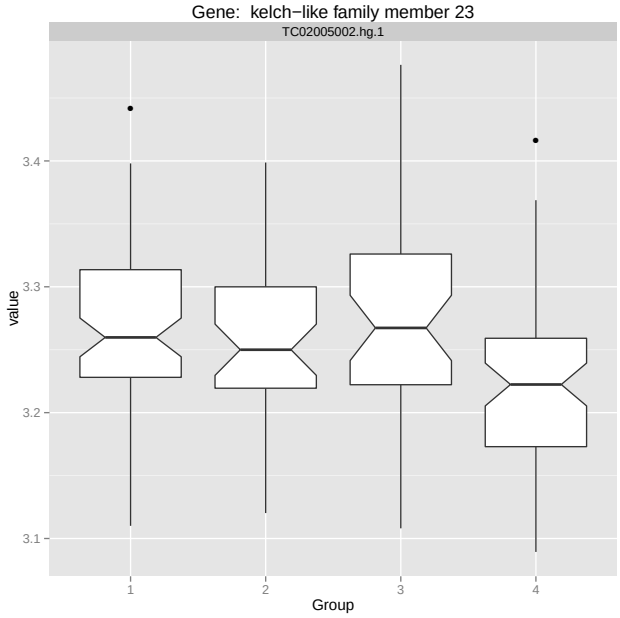


Figure 13: Group Differences for Gene Symbol KLHL23 : kelch-like family member 23

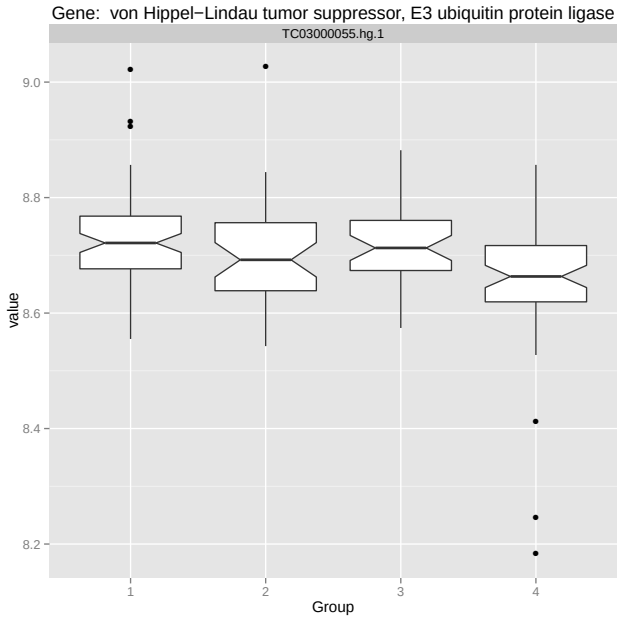


Figure 14: Group Differences for Gene Symbol VHL : von Hippel-Lindau tumor suppressor, E3 ubiquitin protein ligase

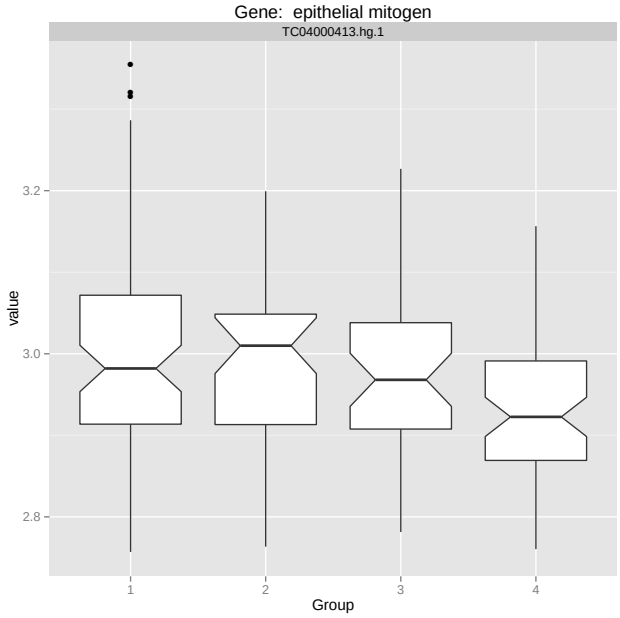


Figure 15: Group Differences for Gene Symbol EPGN : epithelial mitogen

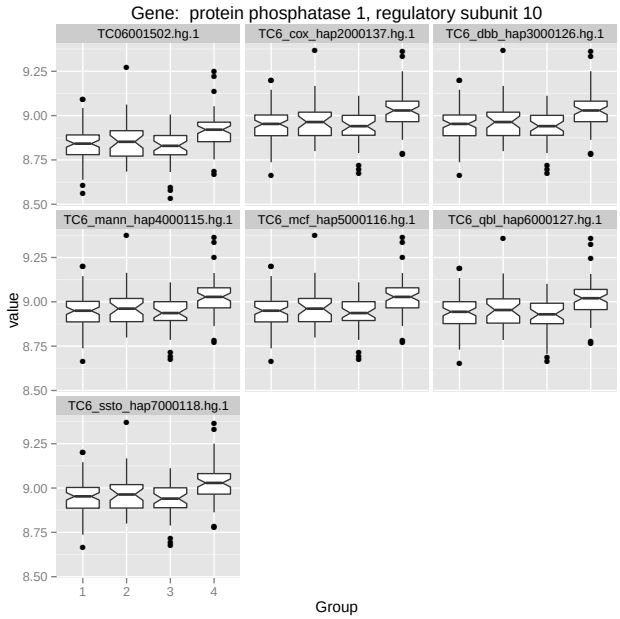


Figure 16: Group Differences for Gene Symbol PPP1R10 : protein phosphatase 1, regulatory subunit 10

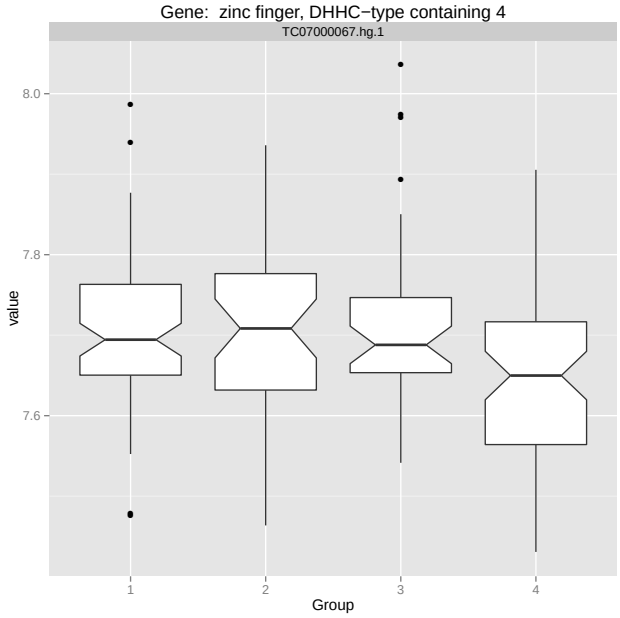


Figure 17: Group Differences for Gene Symbol ZDHHC4 : zinc finger, DHHC-type containing 4

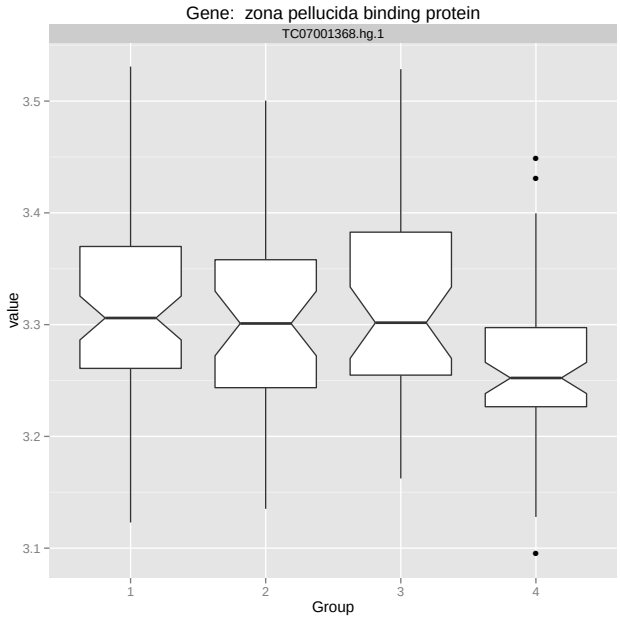


Figure 18: Group Differences for Gene Symbol ZPBP : zona pellucida binding protein

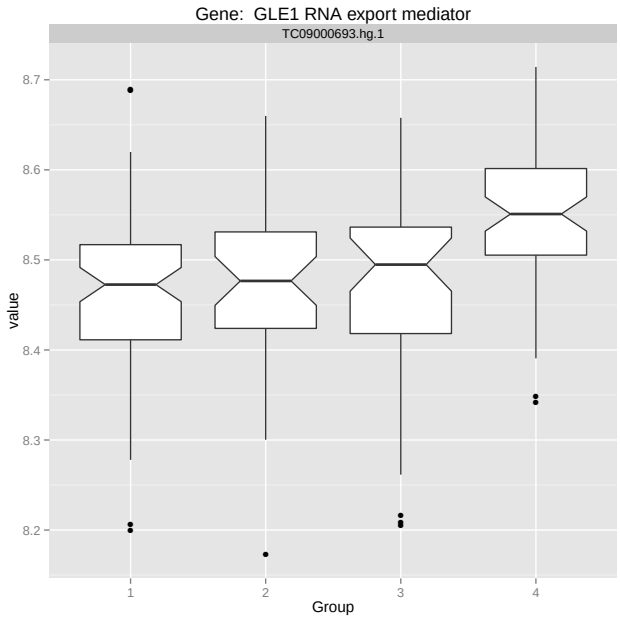


Figure 19: Group Differences for Gene Symbol GLE1 : GLE1 RNA export mediator

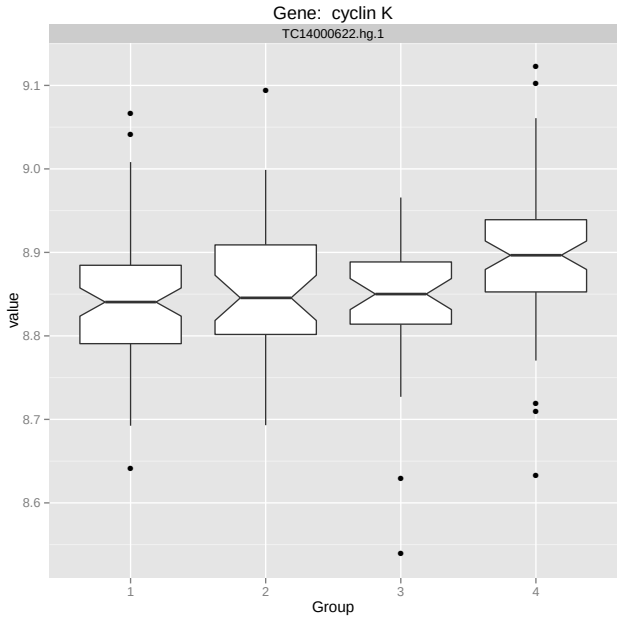


Figure 20: Group Differences for Gene Symbol CCNK : cyclin K

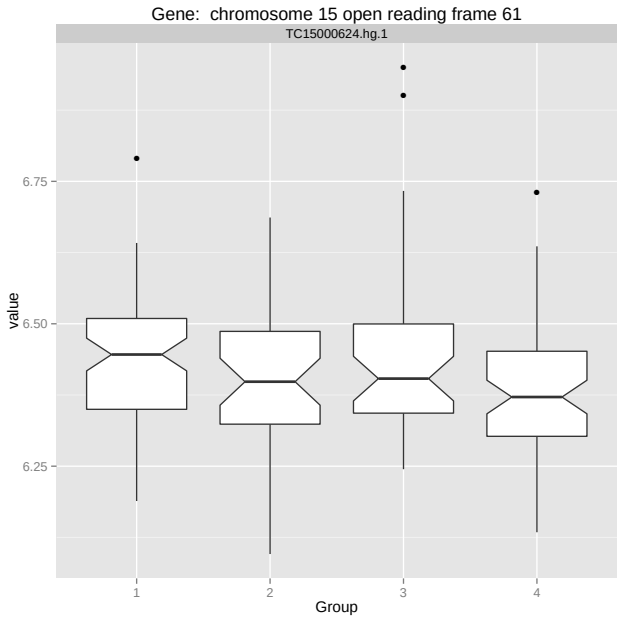


Figure 21: Group Differences for Gene Symbol C15orf61 : chromosome 15 open reading frame 61

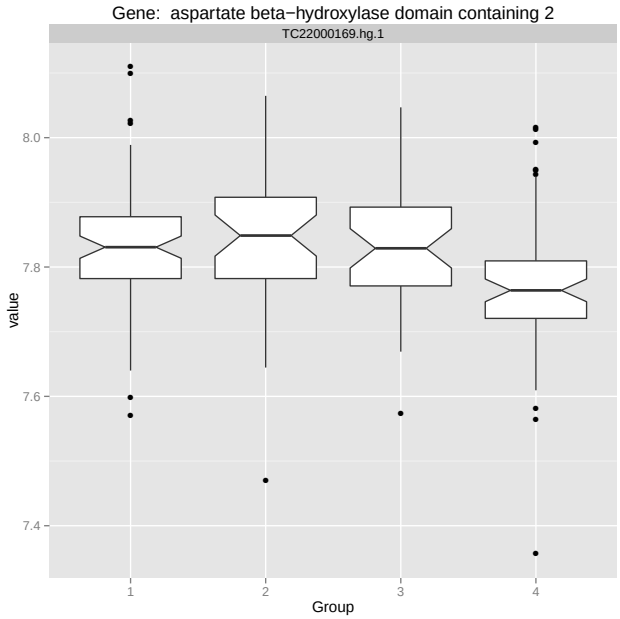


Figure 22: Group Differences for Gene Symbol ASPHD2 : aspartate beta-hydroxylase domain containing 2

Since date scanned is partially confounded with exposure group, the analysis was repeated using only those chips scanned on or after 16thOctober 2013: the earliest date on which exposure group 4 samples were scanned. There were no statistically significant gene expression differences other than for group 1 (control) compared with Group 4 (high exposure). 98 probe sets were statistically significant for the reduced data set.

Each gene was recorded as being statistically significant or not in the full data set, and in the reduced data set. Status in the reduced data set was then cross-tabulated with status in the reduced data set. The results are shown in Table 4. This comparison is made at the level of the gene (Entrez gene ID) rather than at the level of the probesets; some Entrez IDs have more than one probeset.

Full	Reduced Data Set	
Data Set	FALSE	TRUE
FALSE	22091	75
TRUE	154	17

Table 4: Status in the full data set (all dates scanned) vs Status in the Reduced Data Sets (only chips scanned on or after 16thOctober 2010)

Although there is obviously a strong relationship between the two outcomes (Fisher's exact test p value ¡ 10^{-16}), there are 229 genes which are significant in only one of the two data sets. The majority of these divergent results are significant only in the full data set, and not when the early scan dates are excluded. The top 20 probe sets for the reduced data set are shown in Table 5. Again the log fold changes are very small. The maximum effect is a reduction of 1.4 fold.

Comparison	Entrez	Gene Symbol	Gene Name	logFC	AveExpr	t	adj.P.Val
High vs Control	100033438	SNORD116-26	small nucleolar RNA, C/D box 116-26	0.47	7.66	4.24	0.03
High vs Control	100033436	SNORD116-25	small nucleolar RNA, C/D box 116-25	0.42	8.05	4.00	0.04
High vs Control	100033439	SNORD116-27	small nucleolar RNA, C/D box 116-27	0.39	7.29	4.03	0.04
High vs Control	677772	SCARNA6	small Cajal body-specific RNA 6	0.34	10.34	4.27	0.03
High vs Control	692225	SNORD94	small nucleolar RNA, C/D box 94	0.32	10.08	4.08	0.04
High vs Control	677796	SNORA5C	small nucleolar RNA, H/ACA box 5C	0.30	9.14	4.82	0.01
High vs Control	100500815	MIR3687	microRNA 3687	0.30	8.49	4.20	0.03
High vs Control	100158262	SCARNA9L	small Cajal body-specific RNA 9-like	0.29	5.56	4.79	0.01
High vs Control	619383	SCARNA9	small Cajal body-specific RNA 9	0.28	9.22	3.89	0.05
High vs Control	692092	SNORD32B	small nucleolar RNA, C/D box 32B	0.27	6.45	4.91	0.01
High vs Control	692092	SNORD32B	small nucleolar RNA, C/D box 32B	0.27	6.45	4.91	0.01
High vs Control	692092	SNORD32B	small nucleolar RNA, C/D box 32B	0.27	6.45	4.91	0.01
High vs Control	692092	SNORD32B	small nucleolar RNA, C/D box 32B	0.27	6.45	4.91	0.01
High vs Control	692092	SNORD32B	small nucleolar RNA, C/D box 32B	0.27	6.45	4.91	0.01
High vs Control	692092	SNORD32B	small nucleolar RNA, C/D box 32B	0.27	6.45	4.91	0.01
High vs Control	100379295	RNY4P8	RNA, Ro-associated Y4 pseudogene 8	0.26	6.41	3.88	0.05
High vs Control	8968	HIST1H3F	histone cluster 1, H3f	0.26	9.93	3.92	0.05
High vs Control	692092	SNORD32B	small nucleolar RNA, C/D box 32B	0.25	6.44	4.68	0.01
High vs Control	654319	SNORA5A	small nucleolar RNA, H/ACA box 5A	0.25	8.81	4.29	0.03
High vs Control	8360	HIST1H4D	histone cluster 1, H4d	0.25	8.50	4.08	0.04
High vs Control	8339	HIST1H2BG	histone cluster 1, H2bg	0.24	6.79	4.03	0.04
High vs Control	8366	HIST1H4B	histone cluster 1, H4b	0.23	5.65	4.01	0.04
High vs Control	677845	SNORA79	small nucleolar RNA, H/ACA box 79	0.22	6.46	4.02	0.04
High vs Control	100500845	MIR642B	microRNA 642b	0.20	5.75	3.94	0.05
High vs Control	6875	TAF4B	TAF4b RNA polymerase II, TATA box binding protein (TBP)-associated factor, 105kDa	0.20	7.40	3.84	0.05

Table 5: Statistically Significant Differences in Gene Expression by Contrast for data scanned on or after 16th October 2013 ; Results are ranked by absolute $\log_2$ fold change. Gene names and gene symbols are as defined by the HUGO Gene Nomenclature Committee www.genenames.org. Entrez is the NCBI Entrez code for the gene http://www.ncbi.nlm.nih.gov/gene/. Note some gene symbols are associated with multiple probe sets, and therefore have multiple entries in this table.

Comparison	Entrez	Gene Symbol	Gene Name	logFC	AveExpr	t	adj.P.Val
High vs Control	116540	MRPL53	mitochondrial ribosomal protein L53	0.20	8.13	4.73	0.01
High vs Control	8346	HIST1H2BI	histone cluster 1, H2bi	0.19	5.49	3.85	0.05
High vs Control	8969	HIST1H2AG	histone cluster 1, H2ag	0.19	8.74	4.59	0.01
High vs Control	100873822	RNY4P28	RNA, Ro-associated Y4 pseudogene 28	0.19	5.78	4.57	0.01
High vs Control	6628	SNRPB	small nuclear ribonucleoprotein polypeptides B and B1	-0.18	8.03	-4.40	0.03
High vs Control	26823	RNU12-2P	RNA, U12 small nuclear 2, pseudogene	0.18	6.67	4.06	0.04
High vs Control	100616360	MIR2467	microRNA 2467	-0.17	5.31	-3.92	0.05
High vs Control	2197	FAU	Finkel-Biskis-Reilly murine sarcoma virus (FBR-MuSV) ubiquitously expressed	0.17	7.74	4.80	0.01
High vs Control	9937	DCLRE1A	DNA cross-link repair 1A	0.17	6.08	3.90	0.05
High vs Control	100862670	RNY3P7	RNA, Ro-associated Y3 pseudogene 7	0.17	4.05	4.21	0.03
High vs Control	100616355	MIR4531	microRNA 4531	-0.16	7.59	-4.11	0.04
High vs Control	26747	NUFIP1	nuclear fragile X mental retardation protein interacting protein 1	0.16	6.31	4.08	0.04
High vs Control	537	ATP6AP1	ATPase, H+ transporting, lysosomal accessory protein 1	-0.16	9.04	-3.91	0.05
High vs Control	4717	NDUFC1	NADH dehydrogenase (ubiquinone) 1, subcomplex unknown, 1, 6kDa	0.15	6.79	6.05	0.00
High vs Control	8338	HIST2H2AC	histone cluster 2, H2ac	0.15	10.34	3.93	0.05
High vs Control	7675	ZNF121	zinc finger protein 121	0.15	8.37	4.24	0.03
High vs Control	22869	ZNF510	zinc finger protein 510	0.15	6.94	4.10	0.04
High vs Control	9532	BAG2	BCL2-associated athanogene 2	0.15	6.74	3.86	0.05
High vs Control	84316	NAA38	N(alpha)-acetyltransferase 38, NatC auxiliary subunit	0.15	7.58	3.83	0.05
High vs Control	55156	ARMC1	armadillo repeat containing 1	0.15	7.80	4.28	0.03
High vs Control	285237	C3orf38	chromosome 3 open reading frame 38	0.15	8.17	4.13	0.04

Table 5: Statistically Significant Differences in Gene Expression by Contrast for data scanned on or after 16thOctober 2013 ; Results are ranked by absolute $\log_2$ fold change. Gene names and gene symbols are as defined by the HUGO Gene Nomenclature Committee `www.genenames.org`. Entrez is the NCBI Entrez code for the gene `http://www.ncbi.nlm.nih.gov/gene/`. Note some gene symbols are associated with multiple probe sets, and therefore have multiple entries in this table.

Comparison	Entrez	Gene Symbol	Gene Name	logFC	AveExpr	t	adj.P.Val
High vs Control	7112	TMPO	thymopoietin	0.15	8.75	4.56	0.02
High vs Control	6071	RNU6V	RNA, U6 small nuclear variant sequence with SNRPE pseudogene sequence	0.14	7.62	3.84	0.05
High vs Control	128822	CST9	cystatin 9 (testatin)	-0.14	5.80	-4.96	0.01
High vs Control	23788	MTCH2	mitochondrial carrier 2	0.14	8.93	4.10	0.04
High vs Control	5283	PIGH	phosphatidylinositol glycan anchor biosynthesis, class H	0.14	7.00	4.16	0.03
High vs Control	10380	BPNT1	3'(2'), 5'-bisphosphate nucleotidase 1	0.14	7.20	3.92	0.05
High vs Control	10575	CCT4	chaperonin containing TCP1, subunit 4 (delta)	0.14	8.43	3.84	0.05
High vs Control	3716	JAK1	Janus kinase 1	-0.13	10.69	-3.98	0.04
High vs Control	100287948	GM140	uncharacterized LOC100287948	-0.13	3.90	-4.00	0.04
High vs Control	353139	LCE2A	late cornified envelope 2A	-0.13	6.24	-4.11	0.04
High vs Control	9412	MED21	mediator complex subunit 21	0.13	9.40	3.84	0.05
High vs Control	100422895	MIR4294	microRNA 4294	-0.13	6.91	-3.94	0.05
High vs Control	94039	ZNF101	zinc finger protein 101	0.13	9.00	3.82	0.05
High vs Control	7805	LAPTM5	lysosomal protein transmembrane 5	-0.12	12.48	-3.93	0.05
High vs Control	112970	KTI12	KTI12 homolog, chromatin associated (S. cerevisiae)	0.12	7.51	4.01	0.04
High vs Control	2720	GLB1	galactosidase, beta 1	-0.12	8.34	-3.91	0.05
High vs Control	2971	GTF3A	general transcription factor IIIA	0.12	8.80	4.03	0.04
High vs Control	29089	UBE2T	ubiquitin-conjugating enzyme E2T (putative)	0.12	6.28	4.49	0.02
High vs Control	149420	PDIK1L	PDLIM1 interacting kinase 1 like	0.12	7.01	4.17	0.03
High vs Control	55175	KLHL11	kelch-like family member 11	0.12	6.85	4.25	0.03
High vs Control	80067	DCAF17	DDB1 and CUL4 associated factor 17	0.12	8.03	4.02	0.04
High vs Control	693123	MIR449B	microRNA 449b	-0.12	7.45	-3.86	0.05
High vs Control	7596	ZNF45	zinc finger protein 45	0.11	7.13	3.86	0.05

Table 5: Statistically Significant Differences in Gene Expression by Contrast for data scanned on or after 16thOctober 2013 ; Results are ranked by absolute $\log_2$ fold change. Gene names and gene symbols are as defined by the HUGO Gene Nomenclature Committee `www.genenames.org`. Entrez is the NCBI Entrez code for the gene `http://www.ncbi.nlm.nih.gov/gene/`. Note some gene symbols are associated with multiple probe sets, and therefore have multiple entries in this table.

Comparison	Entrez	Gene Symbol	Gene Name	logFC	AveExpr	t	adj.P.Val
High vs Control	100616216	MIR4797	microRNA 4797	0.11	2.64	4.16	0.03
High vs Control	51263	MRPL30	mitochondrial ribosomal protein L30	0.11	8.30	4.19	0.03
High vs Control	400123	LINC00548	long intergenic non-protein coding RNA 548	-0.11	4.70	-4.36	0.03
High vs Control	116236	ABHD15	abhydrolase domain containing 15	0.11	7.72	3.82	0.05
High vs Control	150274	HSCB	HscB mitochondrial iron-sulfur cluster co-chaperone	0.11	6.95	4.88	0.01
High vs Control	100288730	PAN3-AS1	PAN3 antisense RNA 1	0.11	7.50	4.23	0.03
High vs Control	7009	TMBIM6	transmembrane BAX inhibitor motif containing 6	-0.11	11.53	-4.18	0.03
High vs Control	100423018	MIR3156-3	microRNA 3156-3	0.11	4.50	3.85	0.05
High vs Control	5433	POLR2D	polymerase (RNA) II (DNA directed) polypeptide D	0.11	7.05	3.83	0.05
High vs Control	64841	GNPNAT1	glucosamine-phosphate N-acetyltransferase 1	0.11	7.32	4.27	0.03
High vs Control	400743	LOC400743	uncharacterized LOC400743	-0.10	6.33	-3.83	0.05
High vs Control	91942	NDUFAF2	NADH dehydrogenase (ubiquinone) complex I, assembly factor 2	0.10	6.89	3.83	0.05
High vs Control	81191	OR5G5P	olfactory receptor, family 5, subfamily G, member 5 pseudogene	0.10	3.39	3.97	0.04
High vs Control	9798	IST1	increased sodium tolerance 1 homolog (yeast)	-0.10	9.50	-4.10	0.04
High vs Control	6045	RNF2	ring finger protein 2	0.10	8.09	4.00	0.04
High vs Control	4802	NFYC	nuclear transcription factor Y, gamma	-0.10	8.36	-3.91	0.05
High vs Control	5134	PDCD2	programmed cell death 2	0.09	7.89	3.96	0.05
High vs Control	283518	KCNRG	potassium channel regulator	0.09	8.24	4.30	0.03
High vs Control	6988	TCTA	T-cell leukemia translocation altered	0.09	7.59	3.95	0.05

Table 5: Statistically Significant Differences in Gene Expression by Contrast for data scanned on or after 16thOctober 2013 ; Results are ranked by absolute $\log_2$ fold change. Gene names and gene symbols are as defined by the HUGO Gene Nomenclature Committee `www.genenames.org`. Entrez is the NCBI Entrez code for the gene `http://www.ncbi.nlm.nih.gov/gene/`. Note some gene symbols are associated with multiple probe sets, and therefore have multiple entries in this table.

Comparison	Entrez	Gene Symbol	Gene Name	logFC	AveExpr	t	adj.P.Val
High vs Control	114134	SLC2A13	solute carrier family 2 (facilitated glucose transporter), member 13	0.09	6.46	4.15	0.03
High vs Control	7428	VHL	von Hippel-Lindau tumor suppressor, E3 ubiquitin protein ligase	0.09	8.70	4.39	0.03
High vs Control	137695	TMEM68	transmembrane protein 68	0.09	6.29	3.88	0.05
High vs Control	6119	RPA3	replication protein A3, 14kDa	0.09	6.22	4.57	0.01
High vs Control	100101467	ZSCAN30	zinc finger and SCAN domain containing 30	0.08	6.30	3.88	0.05
High vs Control	9125	RQCD1	RCD1 required for cell differentiation1 homolog (S. pombe)	0.08	8.76	3.85	0.05
High vs Control	255324	EPGN	epithelial mitogen	0.08	2.97	3.93	0.05
High vs Control	55146	ZDHHC4	zinc finger, DHHC-type containing 4	0.08	7.69	3.83	0.05
High vs Control	57226	LYRM2	LYR motif containing 2	0.07	6.00	3.95	0.05
High vs Control	84140	FAM161A	family with sequence similarity 161, member A	0.06	4.27	3.91	0.05

Table 5: Statistically Significant Differences in Gene Expression by Contrast for data scanned on or after 16thOctober 2013 ; Results are ranked by absolute $\log_2$ fold change. Gene names and gene symbols are as defined by the HUGO Gene Nomenclature Committee `www.genenames.org`. Entrez is the NCBI Entrez code for the gene `http://www.ncbi.nlm.nih.gov/gene/`. Note some gene symbols are associated with multiple probe sets, and therefore have multiple entries in this table.

Gene expression in the reduced data set was plotted against exposure group for genes associated with the 20 most differential probe sets. These plots are shown in Figures **??** to **??**. Again, differences are small and mostly confined to the high exposure group. The results are, however, more variable because of the reduced sample size.

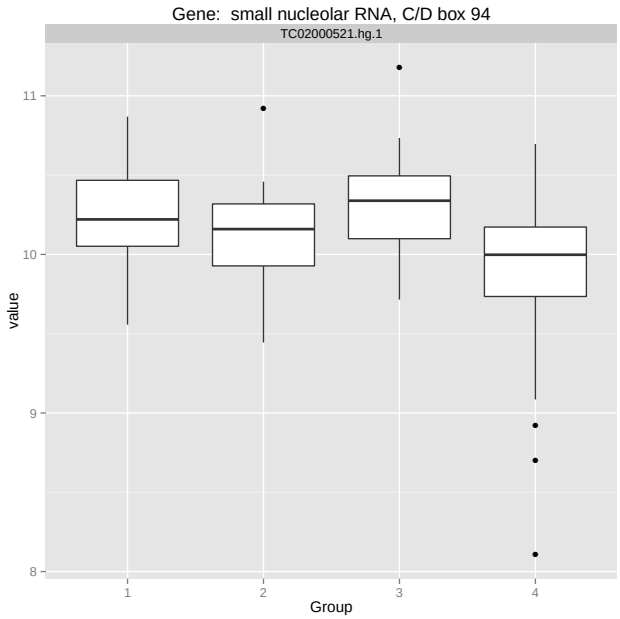


Figure 23: Group Differences for Gene Symbol SNORD94 : small nucleolar RNA, C/D box 94 In Reduced Data Set.

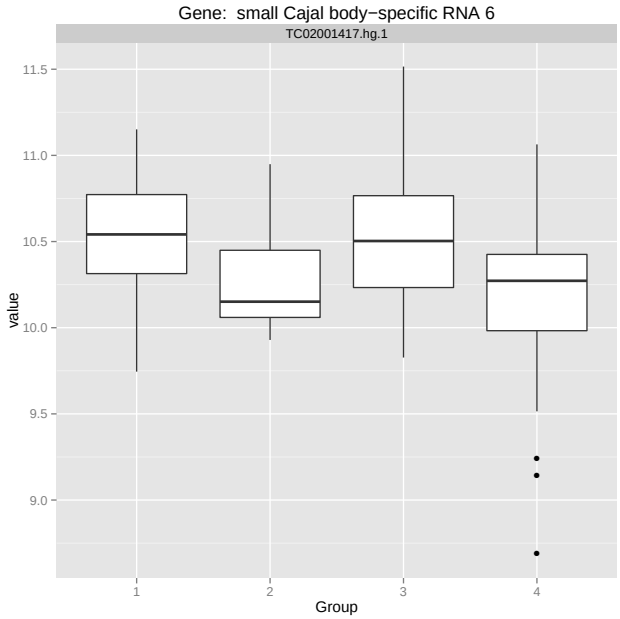


Figure 24: Group Differences for Gene Symbol SCARNA6 : small Cajal body-specific RNA 6 In Reduced Data Set.

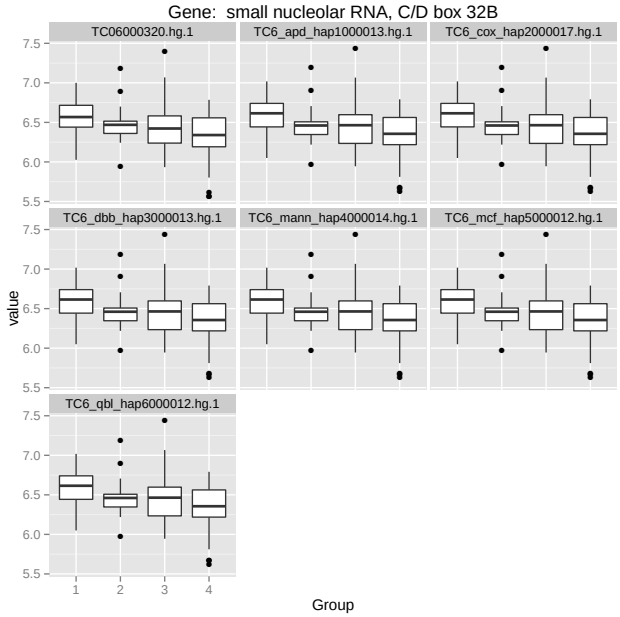


Figure 25: Group Differences for Gene Symbol SNORD32B : small nucleolar RNA, C/D box 32B In Reduced Data Set.

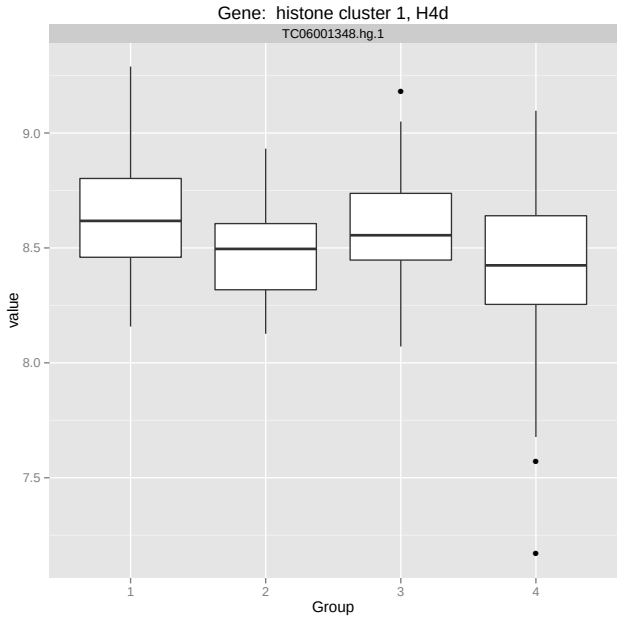


Figure 26: Group Differences for Gene Symbol HIST1H4D : histone cluster 1, H4d In Reduced Data Set.

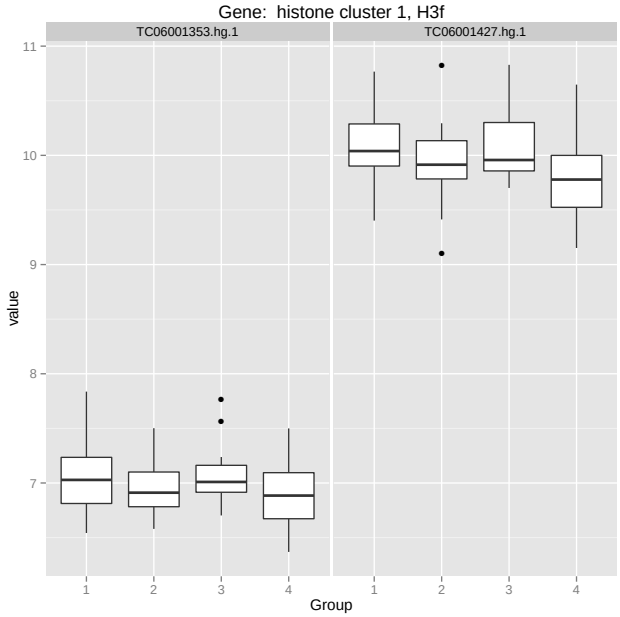


Figure 27: Group Differences for Gene Symbol HIST1H3F : histone cluster 1, H3f In Reduced Data Set.

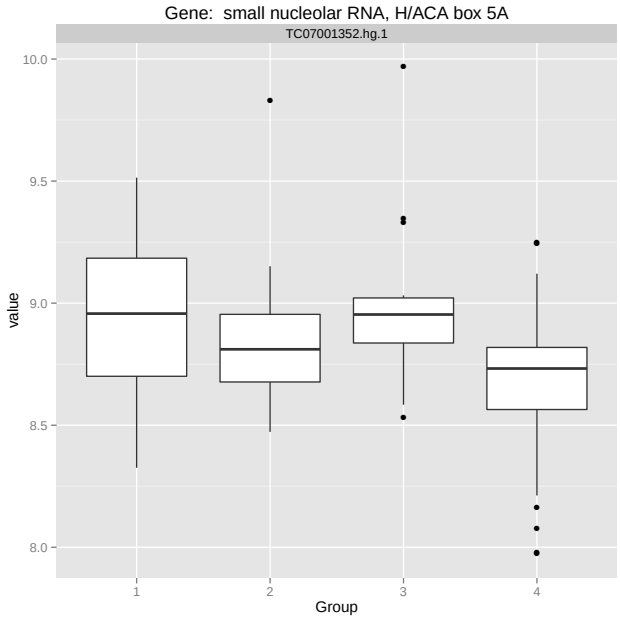


Figure 28: Group Differences for Gene Symbol SNORA5A : small nucleolar RNA, H/ACA box 5A In Reduced Data Set.

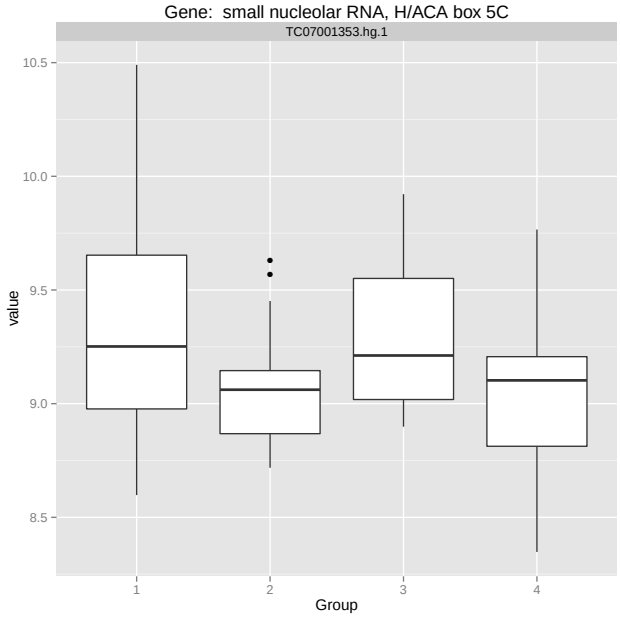


Figure 29: Group Differences for Gene Symbol SNORA5C : small nucleolar RNA, H/ACA box 5C In Reduced Data Set.

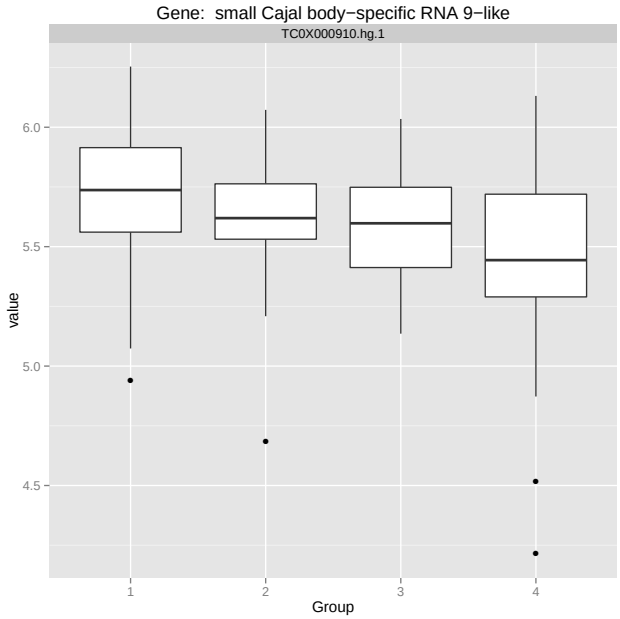


Figure 30: Group Differences for Gene Symbol SCARNA9L : small Cajal body-specific RNA 9-like In Reduced Data Set.

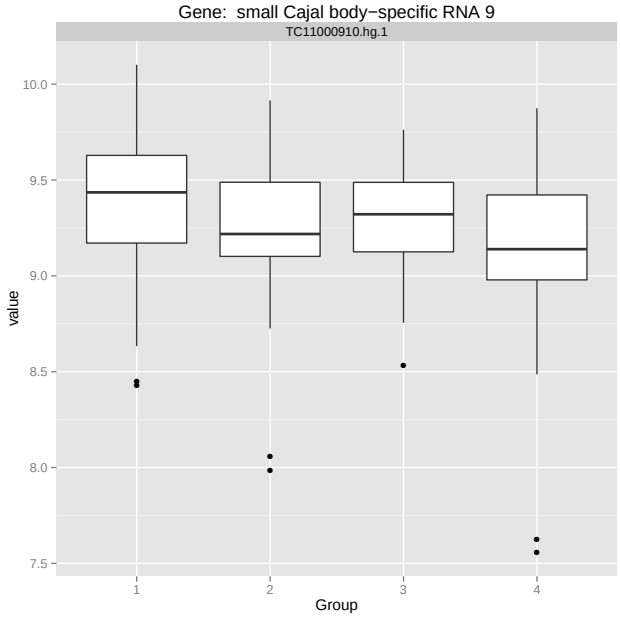


Figure 31: Group Differences for Gene Symbol SCARNA9 : small Cajal body-specific RNA 9 In Reduced Data Set.

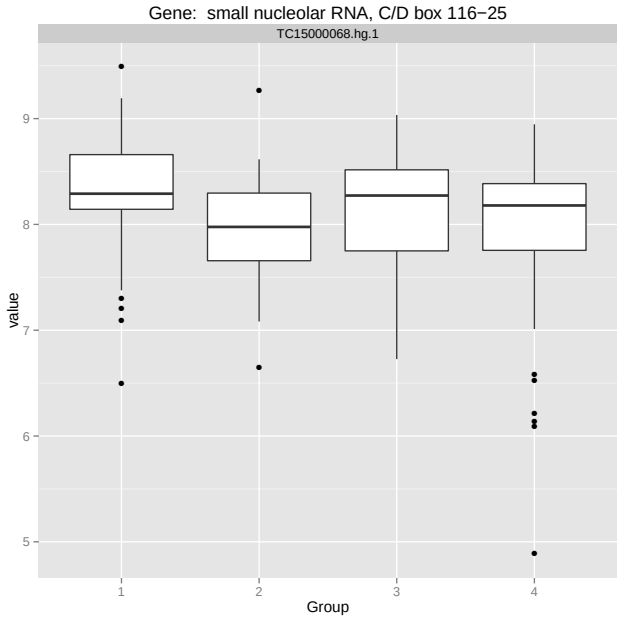


Figure 32: Group Differences for Gene Symbol SNORD116-25 : small nucleolar RNA, C/D box 116-25 In Reduced Data Set.

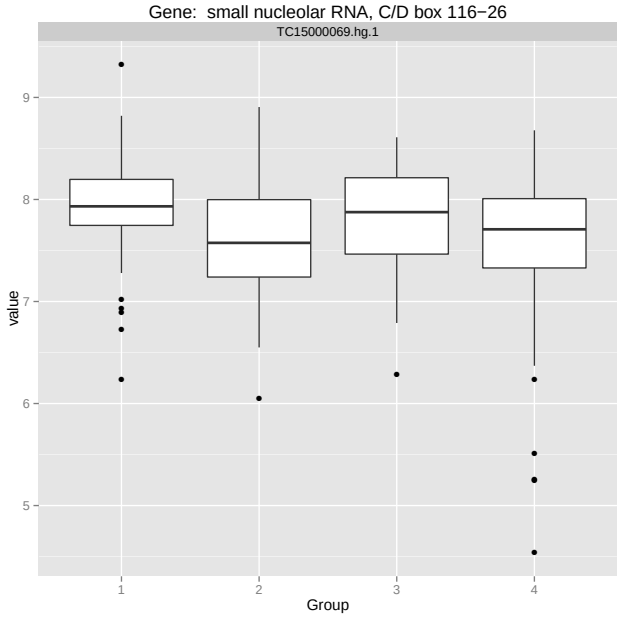


Figure 33: Group Differences for Gene Symbol SNORD116-26 : small nucleolar RNA, C/D box 116-26 In Reduced Data Set.

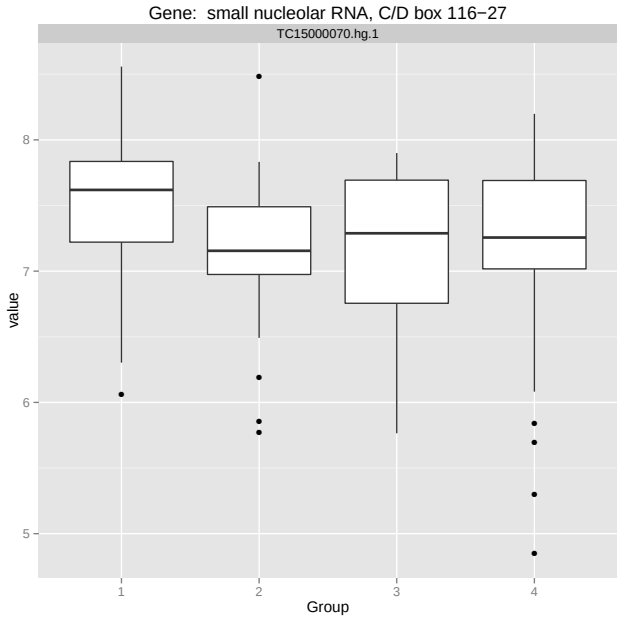


Figure 34: Group Differences for Gene Symbol SNORD116-27 : small nucleolar RNA, C/D box 116-27 In Reduced Data Set.

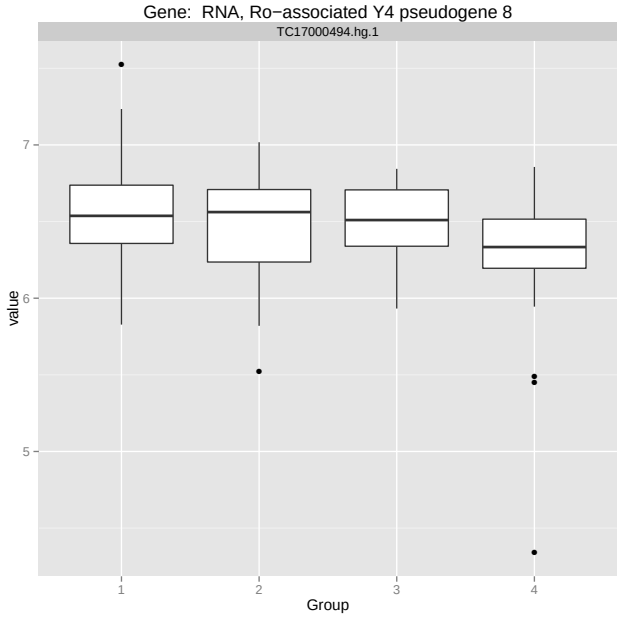


Figure 35: Group Differences for Gene Symbol RNY4P8 : RNA, Ro-associated Y4 pseudogene 8 In Reduced Data Set.

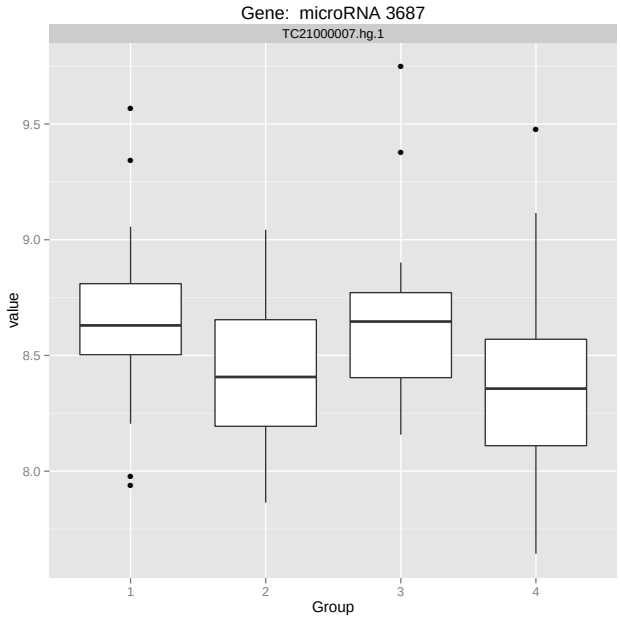


Figure 36: Group Differences for Gene Symbol MIR3687 : microRNA 3687 In Reduced Data Set.

There is evidence of a difference in expression between the control group and the high exposure group. This difference persists, even when chips scanned before October 16th2013 are excluded. There are differences between the two analyses in terms of the genes identified, but this is not unexpected given the relatively arbitrary nature of the selection process.

4.2 Relationship With Health Score

The empirical Bayes analysis was repeated, using health score rather than exposure group. The objective was to identify genes which are correlated with the health score. Following Benjamini Hochberg correction, only 1 probeset showed a statistically significant relationship. That probe set maps to gene symbol TTTY20 (Testis-Specific Transcript, Y-Linked 20 (Non-Protein Coding)). The $\log_2$ fold change associated with a 50 unit change in health score was 0.176: corresponding to a 1.13 fold increase in expression for every 50 unit change. This is a small change in expression for a large change in health score. It is difficult to interpret this effect with any confidence.

Figure 37 shows $\log_2$ gene expression for this gene plotted against health score. The relationship is unimpressive.

Figure 37: Gene Expression for TTTY20 (testis-specific transcript, Y-linked 20 – non-protein coding) vs Health Score

With only a single gene showing change in expression further analysis (e.g. gene ontology enrichment) was infeasible.

5 Ontology Enrichment

It can be difficult to interpret expression changes in a large number of genes. We adopt the approach of Alexa *et al*[3], using hypogeometric tests of enrichment of gene ontology[14] terms in the list of differential genes. The analysis was performed for each of the Gene Ontology consortium domains: molecular function, biological process and cellular component.

Ontology enrichment analysis was performed using the genes statistically significant in the comparison of Group 1 (control) with Group 4 (high exposure). The analysis was performed separately for gene which were over expressed (54 genes) and for genes which were under expressed (64 genes) in the high exposure group. Genes which had low overall expression (median $\log_2$ expression < 2.4) or low overall variability (median absolute deviation from median $\log_2$ expression < 0.064) were filtered out of the analysis.

The statistically significant ontology terms for the over expressed genes in the full data set are shown in table 6. Results are restricted to Gene Ontology terms for which the category size (the maximum number of genes that could be significant for the category) is ≥ 10. Only three molecular function terms are significant; they are associated with mRNA catabolism. Four biological of the five statistically significant Biological Process terms process terms are associated with inter-species interactions. No cellular component terms were statistically significant.

Domain	GO ID	Term	P value	Odds Ratio	Expected Count	Observed Count	Size
BP	GO:0016032	viral process	0.00	6	1	5	646
BP	GO:0044764	multi-organism cellular process	0.00	6	1	5	657
BP	GO:0006281	DNA repair	0.00	7	1	4	402
BP	GO:0044419	interspecies interaction between organisms	0.00	5	1	5	708
BP	GO:0044403	symbiosis, encompassing mutualism through parasitism	0.00	5	1	5	708
MF	GO:0008168	methyltransferase activity	0.00	11	0	3	187
MF	GO:0016741	transferase activity, transferring one-carbon groups	0.00	11	0	3	195
MF	GO:0003824	catalytic activity	0.00	3	8	15	5089

Table 6: Gene Ontology Enrichment for Genes which are Over Expressed in Group 4 vs Group 1. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`.

The statistically significant ontology terms for the under expressed genes in the full data set are shown in table 7. Only terms with category size (the maximum gene count that could be obtained) $\geq$ 10 are tabulated. Many of the biological process terms are associated with viral response. Others are associated with RNA metabolism. The molecular function terms are mostly associated with transcription factor binding. The cellular components are mostly intracellular.

Domain	GO ID	Term	P value	Odds Ratio	Expected Count	Observed Count	Size
BP	GO:0016032	viral process	0.00	6.40	2	9	646
BP	GO:0044764	multi-organism cellular process	0.00	6.28	2	9	657
BP	GO:0044419	interspecies interaction between organisms	0.00	5.80	2	9	708
BP	GO:0044403	symbiosis, encompassing mutualism through parasitism	0.00	5.80	2	9	708
BP	GO:0007219	Notch signaling pathway	0.00	11.58	0	4	143
BP	GO:0006406	mRNA export from nucleus	0.00	18.15	0	3	68
BP	GO:0032922	circadian regulation of gene expression	0.00	51.19	0	2	17
BP	GO:0006405	RNA export from nucleus	0.00	15.93	0	3	77
BP	GO:0046329	negative regulation of JNK cascade	0.00	42.65	0	2	20
BP	GO:0016568	chromatin modification	0.00	5.58	1	6	455
BP	GO:0006366	transcription from RNA polymerase II promoter	0.00	3.39	4	11	1486
BP	GO:0010467	gene expression	0.00	2.63	14	23	5046
BP	GO:0070303	negative regulation of stress-activated protein kinase signaling cascade	0.00	31.98	0	2	26
BP	GO:0032873	negative regulation of stress-activated MAPK cascade	0.00	31.98	0	2	26
BP	GO:0050690	regulation of defense response to virus by virus	0.00	28.42	0	2	29
BP	GO:0051276	chromosome organization	0.00	4.05	2	7	735
BP	GO:0006325	chromatin organization	0.00	4.51	2	6	557
BP	GO:0051028	mRNA transport	0.00	10.31	0	3	117
CC	GO:0005654	nucleoplasm	0.00	5.33	4	15	1398
CC	GO:0070013	intracellular organelle lumen	0.00	4.13	7	20	2634
CC	GO:0044428	nuclear part	0.00	4.19	7	19	2411
CC	GO:0043233	organelle lumen	0.00	4.03	7	20	2691
CC	GO:0031974	membrane-enclosed lumen	0.00	3.94	8	20	2741

Table 7: Gene Ontology Enrichment for for Genes which are Under Expressed in Group 4 vs Group 1. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`.

Domain	GO ID	Term	P value	Odds Ratio	Expected Count	Observed Count	Size
CC	GO:0031981	nuclear lumen	0.00	4.23	6	17	2039
CC	GO:0044424	intracellular part	0.00	15.56	33	44	11954
CC	GO:0005622	intracellular	0.00	14.93	34	44	12081
CC	GO:0044451	nucleoplasm part	0.00	6.41	2	9	614
CC	GO:0044446	intracellular organelle part	0.00	2.95	17	29	6168
CC	GO:0000118	histone deacetylase complex	0.00	24.43	0	3	50
CC	GO:0044422	organelle part	0.00	2.82	18	29	6335
CC	GO:0005643	nuclear pore	0.00	19.12	0	3	63
CC	GO:0005634	nucleus	0.00	2.72	16	27	5758
CC	GO:1902494	catalytic complex	0.00	4.57	2	8	736
CC	GO:0017053	transcriptional repressor complex	0.00	17.38	0	3	69
CC	GO:0046930	pore complex	0.00	15.08	0	3	79
MF	GO:0003682	chromatin binding	0.00	7.01	1	6	357
MF	GO:0005487	nucleocytoplasmic transporter activity	0.00	68.48	0	2	13
MF	GO:0003712	transcription cofactor activity	0.00	5.26	1	6	470
MF	GO:0005515	protein binding	0.00	2.76	22	31	7812
MF	GO:0000989	transcription factor binding transcription factor activity	0.00	4.89	1	6	504
MF	GO:0000988	protein binding transcription factor activity	0.00	4.85	1	6	508
MF	GO:0017124	SH3 domain binding	0.00	10.28	0	3	115

Table 7: Gene Ontology Enrichment for Genes which are Under Expressed in Group 4 vs Group 1. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`.

The enrichment analysis was repeated using genes that were statistically significant in the reduced data set (all chips scanned on or after 16thOctober 2013). The results for over expressed genes are shown in Table **??**. OInly results with a category size (maximum possible number of differential genes) ≥ 10 are reported.

Most of the biological process results for over expressed genes are associated with nucleic acid metabolism and chromatin binding. Most of the molecular function results are associated with chromatin binding.

Domain	GO ID	Term	P value	Odds Ratio	Expected Count	Observed Count	Size
BP	GO:0065004	protein-DNA complex assembly	0.00	12	0	4	134
BP	GO:0071824	protein-DNA complex subunit organization	0.00	11	0	4	157
BP	GO:0044260	cellular macromolecule metabolic process	0.00	3	18	28	6610
BP	GO:0006289	nucleotide-excision repair	0.00	16	0	3	77
BP	GO:0090304	nucleic acid metabolic process	0.00	3	12	21	4360
BP	GO:0006334	nucleosome assembly	0.00	12	0	3	105
BP	GO:0006139	nucleobase-containing compound metabolic process	0.00	2	14	23	5257
BP	GO:0031497	chromatin assembly	0.00	10	0	3	118
CC	GO:0000785	chromatin	0.00	8	1	6	303
CC	GO:0043231	intracellular membrane-bounded organelle	0.00	3	26	37	9298
CC	GO:0043229	intracellular organelle	0.00	4	29	39	10349
CC	GO:0044427	chromosomal part	0.00	5	2	7	542
CC	GO:0000786	nucleosome	0.00	19	0	3	63
CC	GO:1990104	DNA bending complex	0.00	19	0	3	63
CC	GO:0044815	DNA packaging complex	0.00	17	0	3	69
CC	GO:0043227	membrane-bounded organelle	0.00	3	28	38	10134
CC	GO:0005654	nucleoplasm	0.00	3	4	11	1398
CC	GO:0005694	chromosome	0.00	5	2	7	635
CC	GO:0005840	ribosome	0.00	8	1	4	188
CC	GO:0043226	organelle	0.00	3	30	39	10942
CC	GO:0043228	non-membrane-bounded organelle	0.00	2	9	17	3198
CC	GO:0043232	intracellular non-membrane-bounded organelle	0.00	2	9	17	3198
CC	GO:1902494	catalytic complex	0.00	4	2	7	736
CC	GO:0030529	ribonucleoprotein complex	0.00	4	2	6	572
CC	GO:0005634	nucleus	0.00	2	16	25	5758

Table 8: Gene Ontology Enrichment for for Genes which are Over Expressed in Group 4 vs Group 1 for chips scanned on or after October 16th2013. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`.

Domain	GO ID	Term	P value	Odds Ratio	Expected Count	Observed Count	Size
MF	GO:0003677	DNA binding	0.00	3	6	14	2247
MF	GO:0003899	DNA-directed RNA polymerase activity	0.00	22	0	2	38
MF	GO:0034062	RNA polymerase activity	0.00	22	0	2	38

Table 8: Gene Ontology Enrichment for Genes which are Over Expressed in Group 4 vs Group 1, for chips scanned on or after October 16th2013. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`.

Results for under expressed genes are shown in Table 9. Only two terms were statistically significant; both were associated with the vacuolar cellular component.

Domain	GO ID	Term	P value	Odds Ratio	Expected Count	Observed Count	Size
CC	GO:0044437	vacuolar part	0.00	24	0	3	328
CC	GO:0005773	vacuole	0.00	16	0	3	484

Table 9: Gene Ontology Enrichment for Genes which are Under Expressed in Group 4 vs Group 1. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`.

There is no consistency between the ontology terms enriched for the full set of chips and the reduced set of chips.

6 Conclusions

1. There is evidence of differential gene expression between the control group and the high exposure group. This is the case for the full data set, and for the reduced data set including only data from chips scanned on or after 16thOctober 2013.
2. There is some similarity in the set of differential genes for the full data set and for the reduced data set. But these sets are by no means identical.
3. All the gene expression changes are small, especially when compared to those associated with acute insults.
4. There is no consistent picture for gene ontology enrichment, between the full data set and the reduced data set.

Any signals associated with the exposure groups are numerically small, and inconsistent; even if they are statistically significant.

References

[1] Quality Assessment of Exon Arrays Revision Date: 2005-09-27. Whitepaper, Affymetrix Inc., 3420 Central Expressway, Santa Clara, CA 95051, April 2007.

[2] B. Everitt, Sabine Landau, and Morven Leese. *Cluster Analysis*. Edward Arnold, 4 edition, May 2001.

[3] S. Falcon and R. Gentleman. Using GOstats to test gene lists for GO term association. *Bioinformatics*, 23(2):257–8, 2007.

[4] J. A. Hartigan. Consistency of Single Linkage for High-Density Clusters. *Journal of the American Statistical Association*, 76(374):388–394, 1981.

[5] P. Huber. *Robust Statistics*. John Wiley and Sons, 1981.

[6] R. Irizarry, B. Hobbs, F. Collin, Y. Beazer-Barclay, K. Antonellis, U. Scherf, and T. Speed. Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data. *Biostatistics*, 4:249 264, 2003.

[7] I. Jolliffe. *Principal Components Analysis*. Springer-Verlag, 1986.

[8] H. E. Lockstone. Exon array data analysis using Affymetrix power tools and R statistical software. *Briefings in Bioinformatics*, 2011.

[9] I. Lonnstedt and T. Speed. Replicated Microarray Data. *Statistica Sinica*, pages 31–46, 2002.

[10] J. M. Lucas and M. S. Saccucci. Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements. *Technometrics*, 32(1):1–12, 1990.

[11] P. B. Robinson and T. Y. Ho. Average Run Lengths of Geometric Moving Average Charts by Numerical Methods. *Technometrics*, 20(1):85–93, 1978.

[12] B. Silverman. *Density estimation for statistics and data analysis*. Chapman and Hall, London, 1986.

[13] G. K. Smyth. Limma: linear models for microarray data. In R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, and W. Huber, editors, *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, pages 397–420, New York, 2005. Springer.

[14] The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25:25–29, 2000.

[15] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, 57:289–300, 1995.

Appendix 4
Report from Emphron on Bioinformatic analysis of regulatory gene expression array data from leucocytes of exposed workers and controls.

ANALYSIS of Micro RNA ARRAYS

Emphron Informatics Pty Ltd
Author: Mervyn Thomas
Date: 11 June 2014

Jet Fuel and Solvents Exposure Project

Analysis of Micro RNA Arrays

Emphron Informatics Pty Ltd **www.emphron.com**

Report Number: 2012.020
Version Number: 1.0

Client: Jet Fuel and Solvents Exposure Project
Date: July 22, 2014
Author: Mervyn Thomas
Email: mervyn.thomas@emphron.com

Signature: .....

Executive Summary

1. Data were available from 219 high quality micro RNA gene expression assays. Three types of micro RNAs were analysed: stem-loop pre-micro RNAs, mature micro RNAs, and small nucleolar micro RNAs. Stem-Loop pre-micro RNAs and mature micro RNAs were associated with messenger RNA targets using the microRNA.org data base (generated using the miRanda algorithm[8].

2. Exploratory analysis showed that the multivariate disposition of samples in the space of the three types of micro RNA was very similar. There was some evidence of exposure group related differences in the principal components of each type of micro RNA. Mean differences were associated with the lower order principal components, suggesting that whilst these differences were real they were a relatively small part of the total variation. The clinical significance of such differences is unknown.

3. Empirical Bayes analysis of differences between exposure groups revealed clear evidence of differential micro RNA expression for each type of micro RNA. These differences were strongest in comparing group 1 (high exposure) to Group 4 (The controls).

4. Linking stem loop pre micro RNAs and mature micro RNAs to the target mRNAs and conducting ontology enrichment analysis suggested that the major biological processes involved were associated with RNA metabolism and control of gene expression, and neuro-genesis We regard this finding as somewhat speculative, but certainly potentially important.

5. There is reasonable evidence of a difference between controls and the high exposure group in expression of small nucleolar micro RNAs. Given current knowledge it is difficult to ascribe clinical significance to this finding, but impossible to rule out clinical significance. In short, changes in these micro RNAs *are* happening; they may or may not be clinically important.

6. The psychological health score was associated with statistically significant changes in expression of mature micro RNAs. Linking these to mRNAs and conducting ontology enrichment analysis the main biological processes were again control of gene expression and neuro-genesis

7. There is clear evidence of exposure related changes in micro RNA expression, and a modicum of suspicion that these may be associated with psychological health and well being of subjects.

Contents

List of Tables

List of Figures

1 Introduction

1.1 Background and Objectives

The Jet Fuel and Solvents Exposure study has been set up to investigate physiological and other differences between ex Air Force employees who were exposed to aviation fuel and solvents during their service. Subjects were stratified into four exposure groups, where Group 4 is a control group (representing no known exposure) and Group 1 is the highest level of exposure.

Data are also available from other genomic modalities (including cytogentics data and High Throughput mRNA gene expression data) and from proteomic sources. Only the results of micro RNA expression from blood samples, estimated using the Affymetrix micro RNA 2 chip are described in this report.

Each subject was also asked to complete a Health questionnaire (Ref), and data from the psycho social scale of the questionnaire were investigated in conjunction with the gene expression.

1.2 Available Data

Micro RNA gene chip data were available from 224 subjects. Five of these gene chips failed quality assessments and were excluded from analysis (*vide infra*)[1] The break down of subjects with high quality chips by exposure group is shown in Table 1. We note that all group 4 chips were scanned on or after 16thOctober 2013. The comparison of Group 4 with the other groups is therefore partially confounded with date of scanning. This issue is addressed in the Methods section below.

Date	Exposure Group			
Scanned	1	2	3	4
2013-01-25	11	6	7	0
2013-07-18	8	8	4	0
2013-07-24	15	9	8	0
2013-10-02	8	3	6	13
2013-10-10	9	1	10	11
2013-11-22	15	7	2	5
2014-01-01	10	3	5	5
2014-01-03	5	2	0	23

Table 1: Samples By Exposure Group and Date Scanned

2 Pre Processing of Gene Chip Data

2.1 Expression Estimates

Raw chip data were processed to give estimates of micro RNA expression using the RMA algorithm[7], as implemented in the Affymetrix Power Tools Suite[13]. The Affymetrix micro RNA chip contains sequences from a number of species, and of a number of types. Only human sequence probe sets were retained for analysis, and data sets were generated for mature micro RNAs (20 nucleotides), stem-loop pre-micro RNAs (70 nucleotides) and small nucleolar RNAs.

Throughout the remainder of this report it should be remembered that stem-loop pre-micro RNAs act as precursors to mature micro RNAs. They are cleaved by the 'dicer' enzyme to produce the mature micro RNA. Mature micro RNAs bind with complementary sequences on mRNAs, and either inhibit translation, silence translation, or promote mRNA degradation. High levels of expression of stem-loop pre-micro RNAs are therefore associated with high levels of micro RNAs of matching species, and decreased translation or abundance of the matching mRNAs.

Small nucleolar RNAs are mostly associated with the methylation and pseudouridylation of ribosomal RNAs, but they can also interact with mRNAs[19].

[1]This is actually rather better than standard performance for a genomics laboratory, and should not be considered as a criticism of laboratory processing of these complex assays.

Affymetrix micro RNA annotation is inconsistent and incomplete. Small nucleolar RNAs were considered to be those probesets with sequence type ”HAcaBox”, ”CDBox” or snoRNA”. Affymetrix annotate probe sets with sequence type ”snoRNA” with Ensembl IDs, whereas those with type ”HAcaBox” or ”CDBox” are annotated with the gene ID as defined in the miRBase database[4] accessed via `http://www.mirbase.org/`. Mature micro RNAs and stem-loop micro RNAs are annotated with the miRBase transcript name. Annotations for sequence type ”snoRNA” were converted from Ensembl IDs to external gene names using Ensembl Biomart[10].

Relatively good target information exists for pre-micro RNAs and for mature micro RNAs[2]. Target relationships were obtained from the microRNA.org database[2] accessed at `http://www.microrna.org`. Target relationships that were characterised by strong complementarity and good cross species conservation [2, page D150] were retained for analysis. Good conservation was defined by a conservation score of 0.57 or more: which is approximately the cut off for conservation within mammals.

2.2 Quality Control

Standard quality metrics were generated as recommended by Affymetrix[1].

One of those quality metrics were selected for further examination: ‘all+probesets_mad_residual_mean’ This is the mean of the absolute deviation of the residuals from the median. The RMA algorithm produces a model for probe level responses, based on the particular probe and the probe set to which it belongs. Residuals from this model represent departures from the model prediction for the given probe on the given chip. The mean absolute deviation for a chip provides a measure of how well the particular chip fits the RMA model: which was developed across all chips. A high mean absolute deviation of the residuals from the median suggests quality problems for the chip concerned.

Chips were sorted by date and time of scanning, and an exponentially weighted moving average control (EWMA) charts[15, 16] was generated for the metric. The EWMA chart is a powerful method of identifying when a process moved out of control. It is not intended to identify individual extreme values; the EWMA is more sensitive to runs of extreme values. It generally lags the onset of extreme values by a few observations.

Histograms were constructed for the metric, and robust Huber estimates[6] of the mean and standard deviation were used to generate mean $\pm$ 3 standard deviation ranges.

Figure 1 shows the EWMA plot for mean absolute deviation of residuals. There is a period towards the end of the sequence in which the process appears to be out of control. Figure 2 shows a histogram and kernel density estimate[17] of the distribution of this metric. The distribution is relatively well behaved, with a small number of extreme values. These extreme values occur during the period highlighted on the EWMA chart.

Inspection of the data revealed that these extreme values were for five RNA samples obtained by manual extraction. The samples were dropped from down stream analysis.

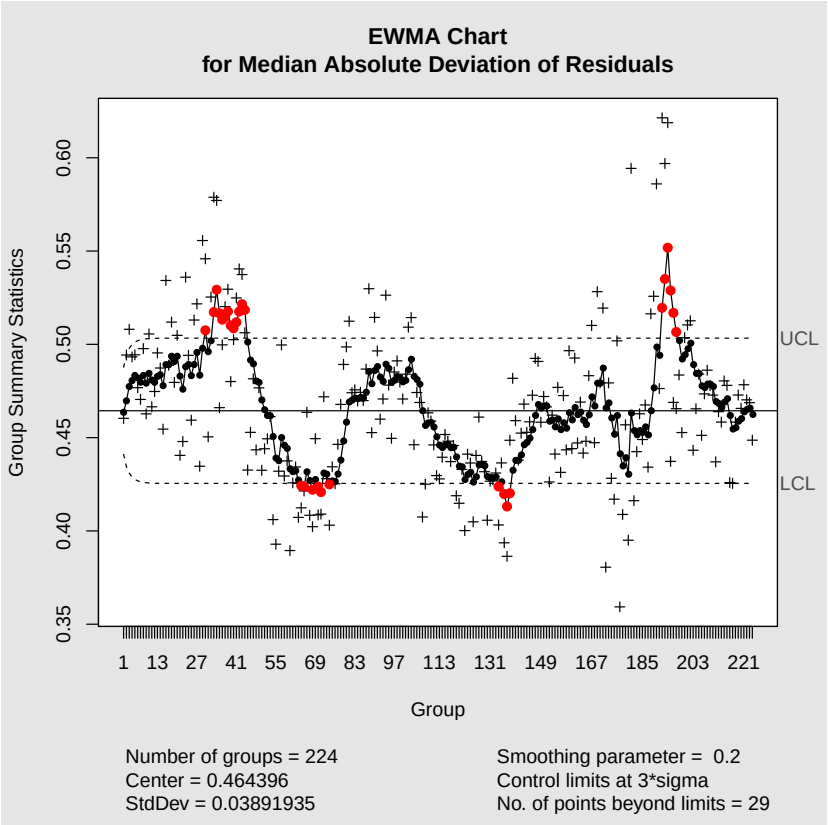


Figure 1: Exponentially Weighted Moving Average (EWMA) Control Chart for Mean Absolute Deviation of Residuals

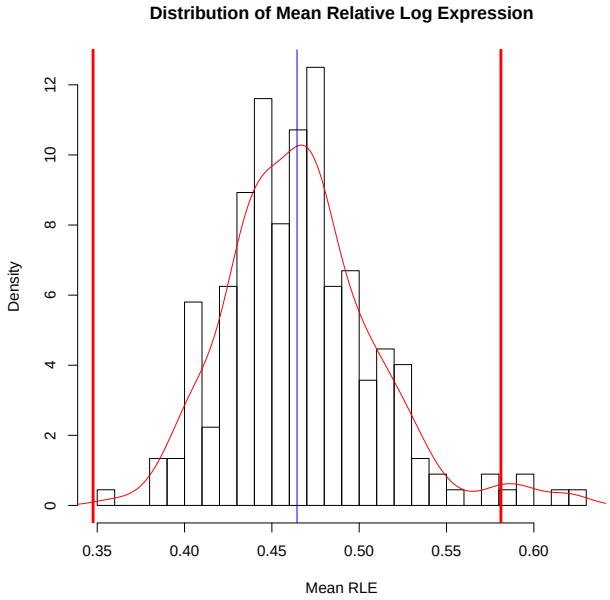


Figure 2: Distribution of Mean Absolute Deviation of Residuals

The quality metrics calculated for these data were very satisfactory. A small number of extreme values were associated with documented differences in chip processing. These chips have been excluded from further analysis.

Probe sets which had low overall expression (median across all chips $\log_2$ expression < 2.4) or low overall variability (median absolute deviation from median across all chips $\log_2$ expression < 0.064) were filtered out of all further analysis analysis.

3 Exploratory Analysis of Gene Expression

3.1 Exploratory Methods

Exploratory analysis was conducted separately for each of the sequence types: mature micro RNAs, stem-loop pre-micro RNAs and small nucleolar micro RNAs. Principal components[9] of the gene expression matrix (represented as samples X genes) were calculated. The scree plots were generated, showing the variance of each principal component,m and each sample was plotted in the space of then first two principal components. Observations were colour coded by exposure group. Box and whisker plots were generated showing the distribution of the first eight principal components by group, and a between groups analysis of variance was conducted for the first eight principal components.

The gene expression data were clustered using three hierarchical clustering algorithms[3]: complete, average and single linkage[5]. Cluster plots are often used to over-interpret data. Apparent cluster structure is frequently illusory. Cluster solutions should only be interpreted when they are markedly consistent across different algorithms.

3.2 Mature Micro RNAs

Figure 3 shows the scree plot for the principal components analysis: giving the variance for each of the first ten principal components. The first principal component dominates, representing 4 times the variance of the second component.

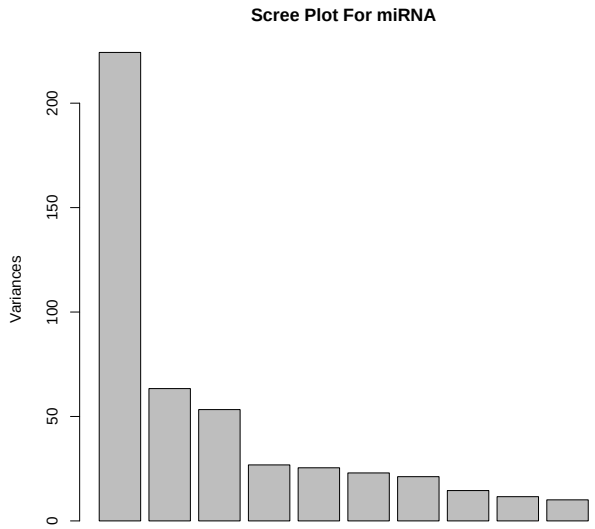


Figure 3: Variance By Principal Component For Mature Micro RNAs

Figure 4 shows the scatter plot for the first two components of the mature micro RNA data. The majority of samples have first component scores close to zero, but there is a long tail of samples with large and negative values. There is some suggestion that more of the exposed samples are in this tail than the control samples.

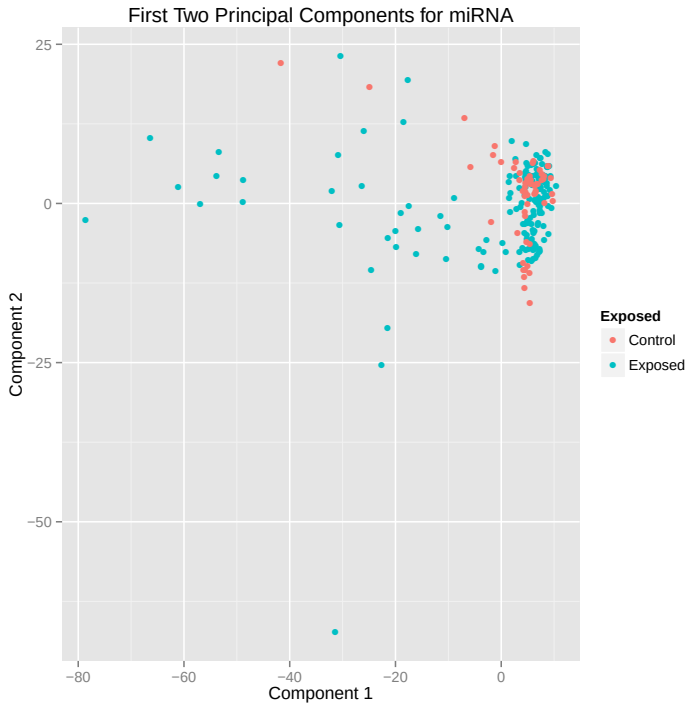


Figure 4: Scatter Plot For First Two Components of mature Micro RNAs

Figure 5 shows box and whisker plots of the first nine principal components by exposure group. There is a suggestion of between group differences for components 5, 7, 8 and 9. These are associated with exposure group (the controls) for components 5, 7 and 9, and with group 1 (the highest exposure group) for component 8.

The analysis of variance for each component is summarised by Table 2. This conforms the existence of statistically significant between group differences for components 5, 7, 8 and 9. This analysis suggests that there may be differences between groups in mature micro RNAs, mostly associated with group 4. These differences, however, are in relatively low order principal components and do not represent a large contribution ton total variation in micro RNA expression.

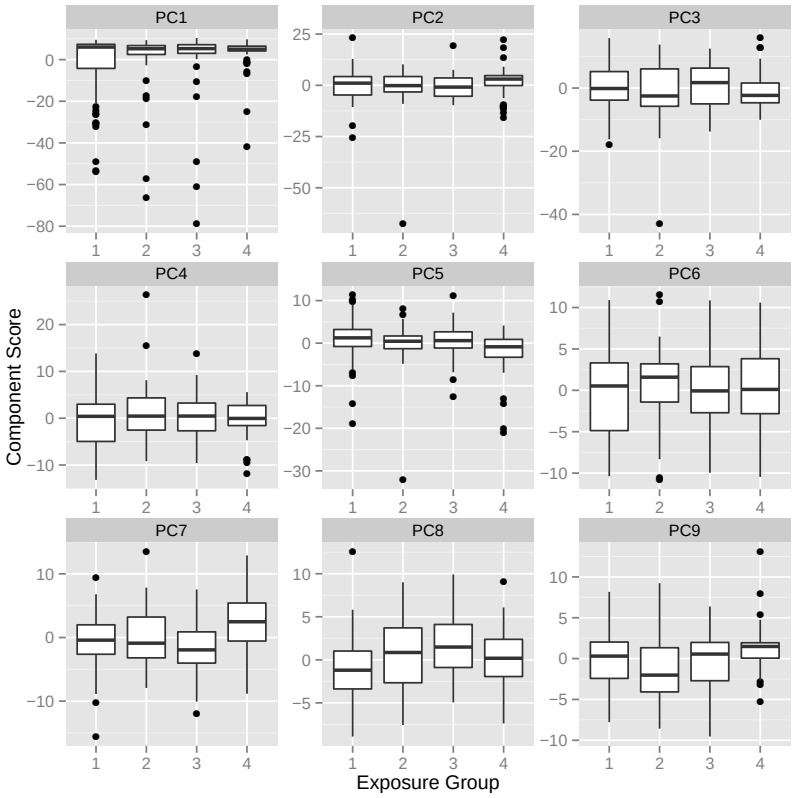


Figure 5: Box And Whisker Plots By Exposure Group For Principal Components of mature Micro RNAs

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
PC 1	3	1032.81	344.27	1.55	0.2035
PC 2	3	244.19	81.40	1.29	0.2788
PC 3	3	148.90	49.63	0.93	0.4269
PC 4	3	140.02	46.67	1.76	0.1560
PC 5	3	306.55	102.18	4.19	0.0066
PC 6	3	24.38	8.13	0.35	0.7887
PC 7	3	479.91	159.97	8.31	0.0000
PC 8	3	207.17	69.06	5.01	0.0022
PC 9	3	181.14	60.38	5.53	0.0011

Table 2: Analysis of Variance For Principal Components of miRNA

Single, average and complete linkage cluster dendrograms are shown in Figures 6, 7 and 8 respectively. Visually the dendrograms are quite different. Closer examination reveals that they are similar only in so far as their major structure is associated with a few extreme samples, rather than the exposure groups. That is, samples occurring as isolated nodes or in small groups in one solution tend to do so in all solutions. Those samples are JF1072, JF2024, JF2033, JF2068 and JF3020.

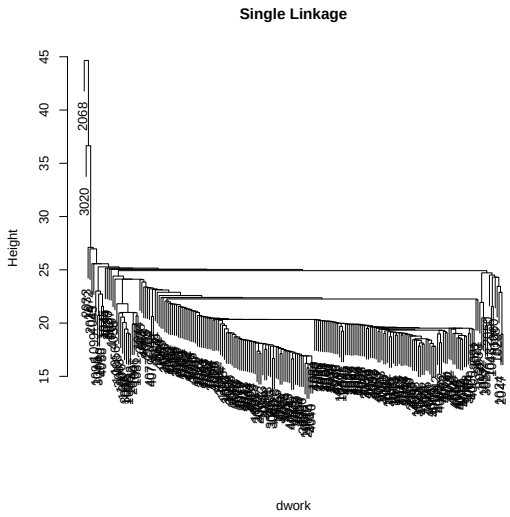


Figure 6: Single Linkage Cluster Analysis For Mature Micro RNAs

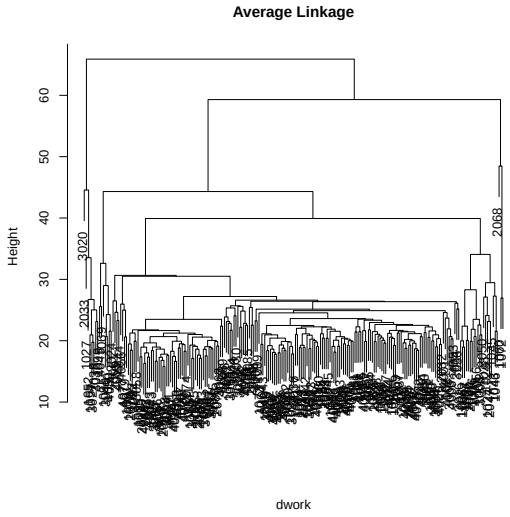


Figure 7: Average Linkage Cluster Analysis For Mature Micro RNAs

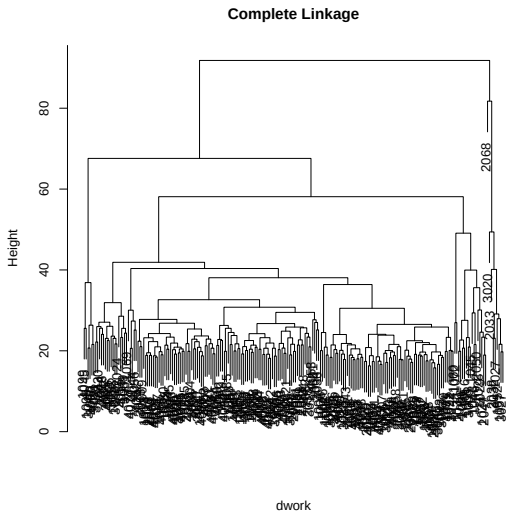


Figure 8: Complete Linkage Cluster Analysis For Mature Micro RNAs

3.3 Stem-Loop Pre Micro RNAs

Figure 9 shows the scree plot for the principal components analysis: giving the variance for each of the first ten principal components. The first principal component dominates, representing 3 times the variance of the second component.

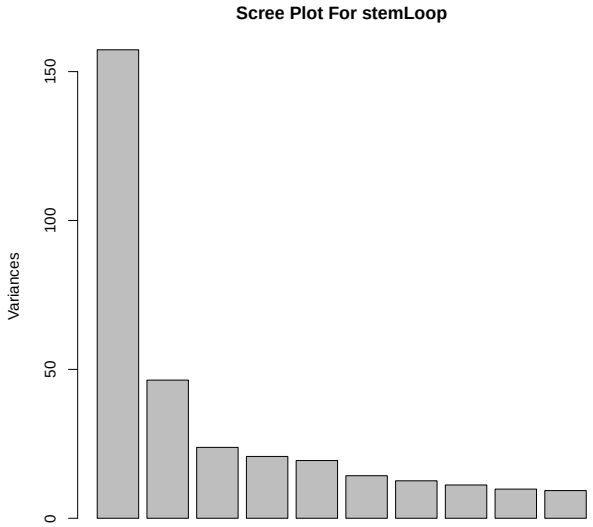


Figure 9: Variance By Principal Component For Stem Loop pre-Micro RNAs

Figure 10 shows the scatter plot for the first two components of the Stem Loop pre-micro RNA data. The majority of samples have first and second component scores close to zero, but there is a long tail of samples with large and negative values on component 1. Most of these samples are from the exposed groups. This plot is markedly similar to that seen for mature micro RNAs, and the correlation between the first principal component scores for mature micro RNAs and stem loop pre-micro RNAs is 0.99. The correlation coefficient for the second component scores is 0.78. There are very strong similarities between the multivariate disposition of samples in the space of the stem loop pre-micro RNAs and the space of the micro RNAs.

It would be possible to establish this more powerfully with a generalised procrustes analysis[12, Chapter 5], but the point is well demonstrated by these simple correlations.

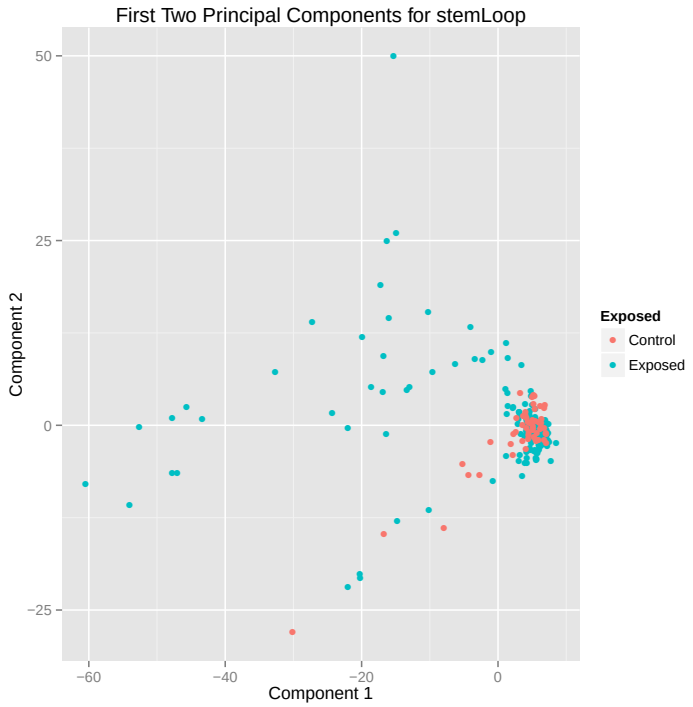


Figure 10: Scatter Plot For First Two Components of Stem Loop pre-Micro RNAs

Figure 11 shows box and whisker plots of the first nine principal components by exposure group. There is a suggestion of between group differences for components 4, 6 and 8. These are mostly associated with exposure group 4, but there is some suggestion of a trend across increasing exposure groups for component 8.

The analysis of variance for each component is summarised by Table 3. This conforms the existence of statistically significant between group differences for components 3, 4, 6 and 8. This analysis suggests that there may be differences between groups in Stem Loop pre-micro RNAs, mostly associated with differences between the control group and the exposed groups. These differences, however, are in relatively low order principal components and do not represent a large contribution to total variation in micro RNA expression.

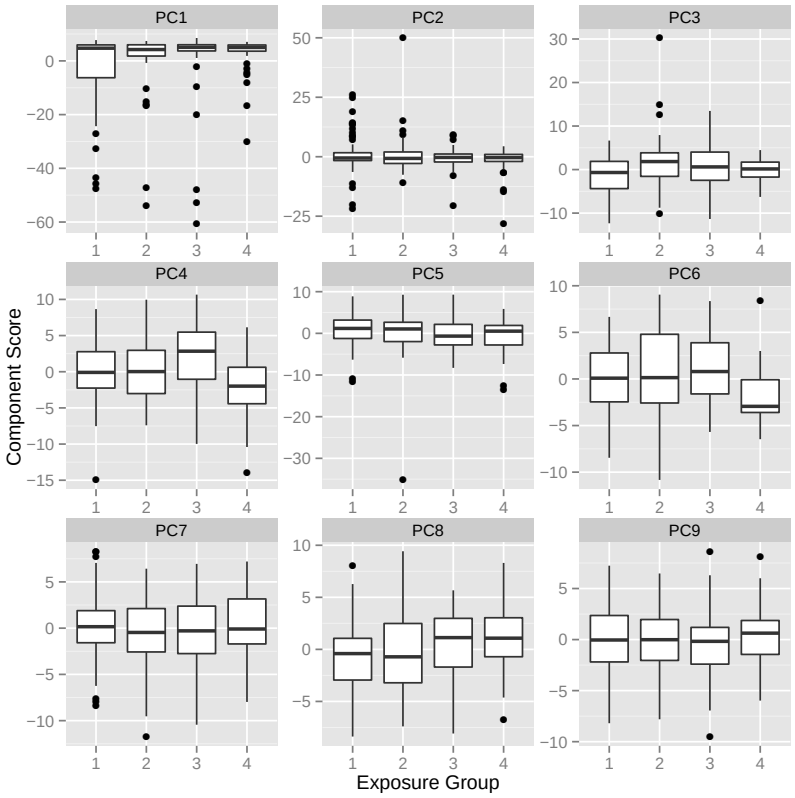


Figure 11: Box And Whisker Plots By Exposure Group For Principal Components of Stem Loop pre-Micro RNAs

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
PC 1	3	930.97	310.32	2.00	0.1151
PC 2	3	201.70	67.23	1.46	0.2269
PC 3	3	346.82	115.61	5.13	0.0019
PC 4	3	464.73	154.91	8.20	0.0000
PC 5	3	64.56	21.52	1.11	0.3453
PC 6	3	322.95	107.65	8.29	0.0000
PC 7	3	33.07	11.02	0.87	0.4550
PC 8	3	132.16	44.05	4.11	0.0073
PC 9	3	17.73	5.91	0.60	0.6154

Table 3: Analysis of Variance For Principal Components of Stem Loop pre-Micro RNAs

Single, average and complete linkage cluster dendrograms are shown in Figures 12, 13 and 14 respectively. Visually the dendrograms are quite different. Closer examination reveals that they are similar only in so far as their major structure is associated with a few extreme samples, rather than the exposure groups. That is, samples occurring as isolated nodes or in small groups in one solution tend to do so in all solutions. Those samples are JF1072, JF2068 and JF3020. These samples also formed isolated clusters for the mature micro RNAs.

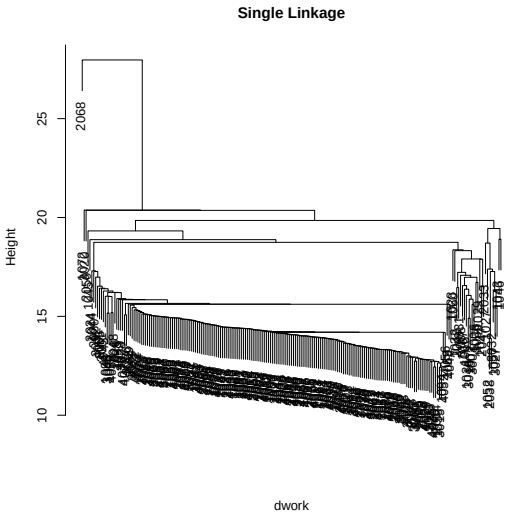


Figure 12: Single Linkage Cluster Analysis For Mature Micro RNAs

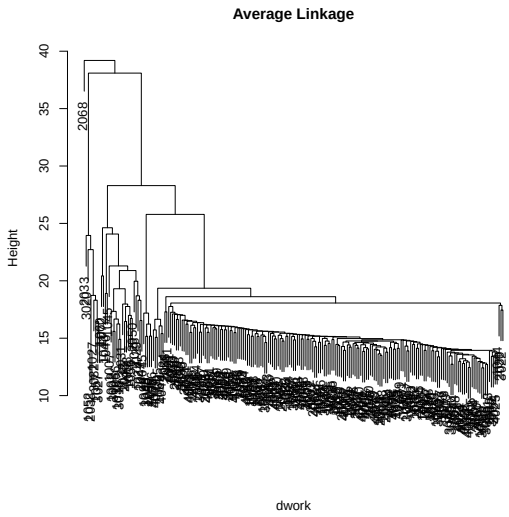


Figure 13: Average Linkage Cluster Analysis For Mature Micro RNAs

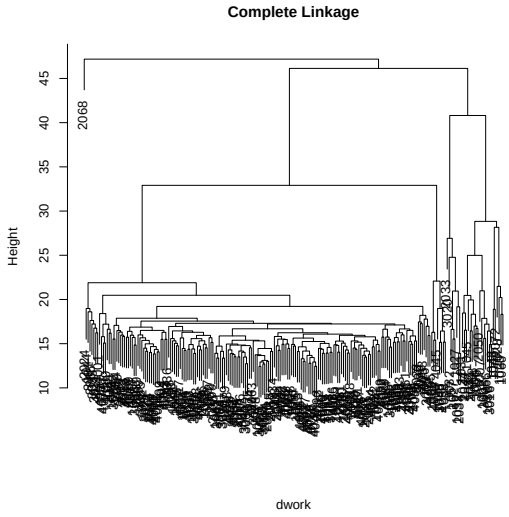


Figure 14: Complete Linkage Cluster Analysis For Mature Micro RNAs

3.4 Small Nucleolar Micro RNAs

Figure 15 shows the scree plot for the principal components analysis: giving the variance for each of the first ten principal components. The first principal component dominates, representing 4 times the variance of the second component.

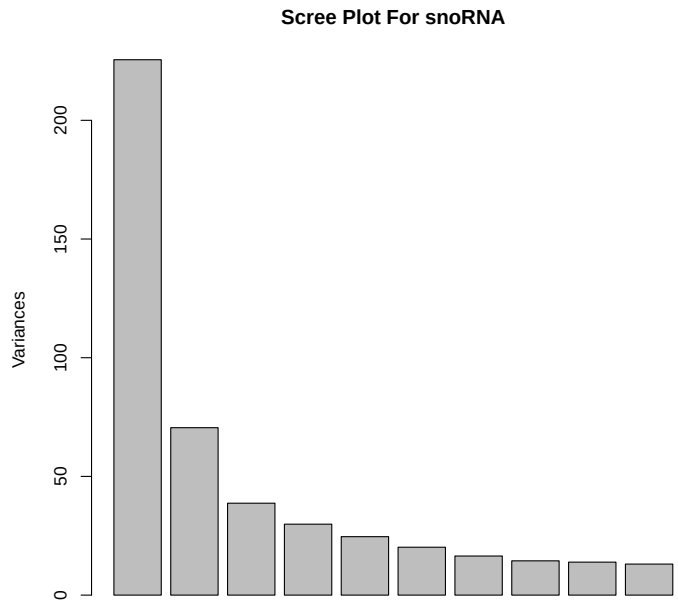


Figure 15: Variance By Principal Component For Small Nucleolar Micro RNAs

Figure 16 shows the scatter plot for the first two components of the small nucleolar micro RNA data. The majority of samples have first component scores close to zero, but there is a long tail of samples with large and negative values. Most of the samples in this long tail are from subjects in the exposed groups. This pattern is very similar to that seen for the mature micro RNAs and for the stem-loop micro RNAs. The Pearson correlation coefficient between the first component score for mature micro RNAs and small nucleolar micro RNAs is 0.98, and the correlation between scores on the first component for stem-loop micro RNAs and small nucleolar RNAs is 0.99. Correlations on the second component are 0.88 and 0.93.

That is, the multivariate disposition of samples in the space of the small nucleolar RNAs is very similar to the disposition in the space of the stem loop micro RNAs and the mature micro RNAs. The question. of course, arises as to whether this reflects common underlying biological processes or common artefacts in laboratory processing.

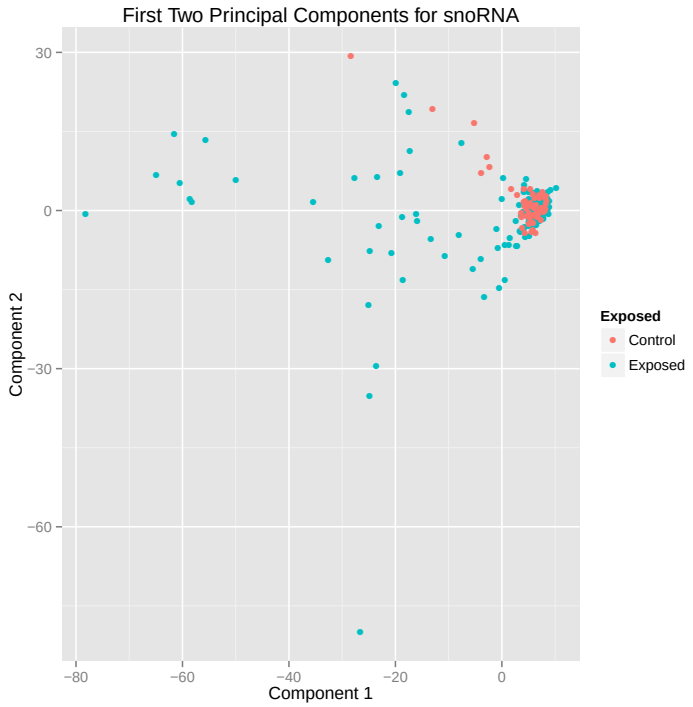


Figure 16: Scatter Plot For First Two Components of Small Nucleolar Micro RNAs

Figure 17 shows box and whisker plots of the first nine principal components by exposure group. There is a suggestion of between group differences for components 3, 4 and 6. These are mostly associated with the control group for components 4 and 6, but the high exposure group for component 3.

The analysis of variance for each component is summarised by Table 4. This conforms the existence of statistically significant between group differences for components 3, 4 and 6. This analysis suggests that there may be differences between groups in small nucleolar micro RNAs, mostly associated with differences between the control group and exposed groups. These differences, however, are not in the dominant principal components.

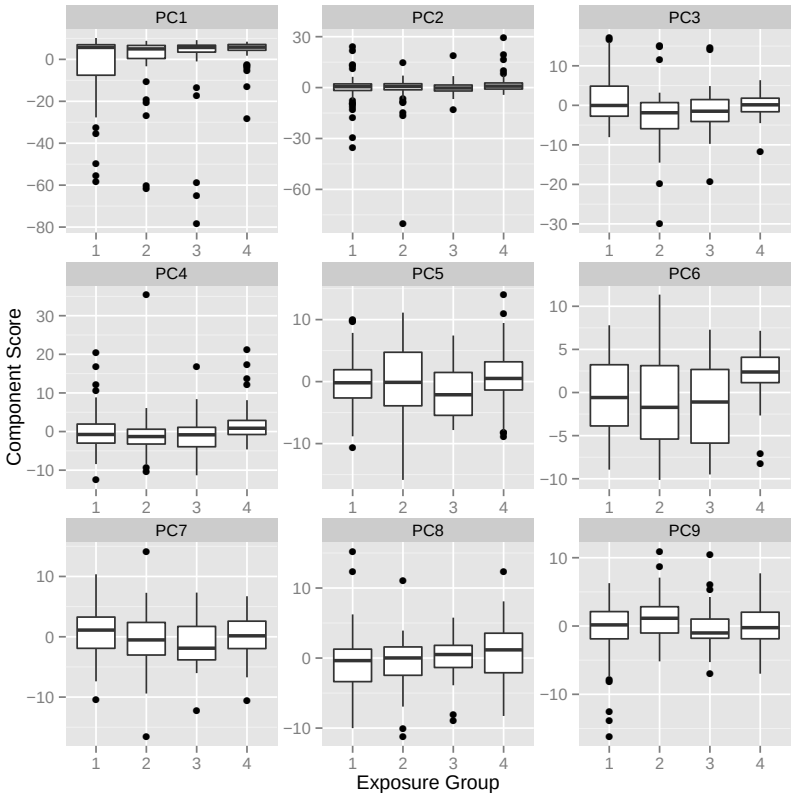


Figure 17: Box And Whisker Plots By Exposure Group For Principal Components of mature Micro RNAs

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
PC 1	3	1595.89	531.96	2.40	0.0684
PC 2	3	413.08	137.69	1.98	0.1182
PC 3	3	582.57	194.19	5.31	0.0015
PC 4	3	310.72	103.57	3.59	0.0146
PC 5	3	132.31	44.10	1.81	0.1461
PC 6	3	402.11	134.04	7.21	0.0001
PC 7	3	92.47	30.82	1.89	0.1315
PC 8	3	102.43	34.14	2.41	0.0679
PC 9	3	86.60	28.87	2.11	0.1000

Table 4: Analysis of Variance For Principal Components of snoRNA

Single, average and complete linkage cluster dendrograms are shown in Figures 18, 19 and 20 respectively. Visually the dendrograms are quite different. Closer examination reveals that they are similar only in so far as their major structure is associated with a few extreme samples, rather than the exposure groups. That is, samples occurring as isolated nodes or in small groups in one solution tend to don son in all solutions. Those samples are JF2033, JF2068 and JF3020; these samples were also in isolated clusters for mature micro RNAs.

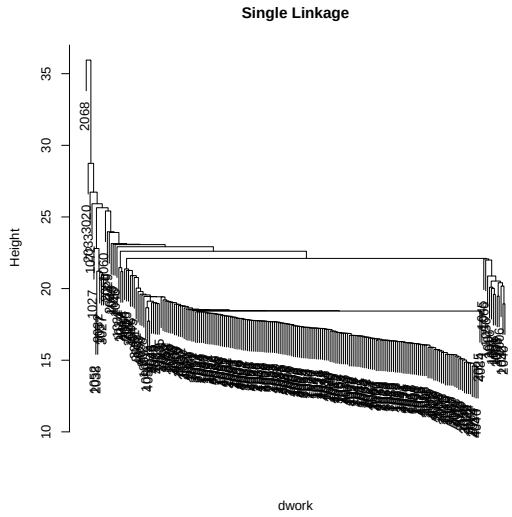


Figure 18: Single Linkage Cluster Analysis For Mature Micro RNAs

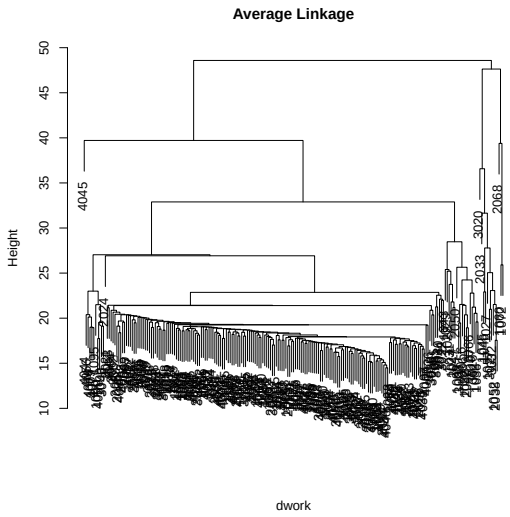


Figure 19: Average Linkage Cluster Analysis For Mature Micro RNAs

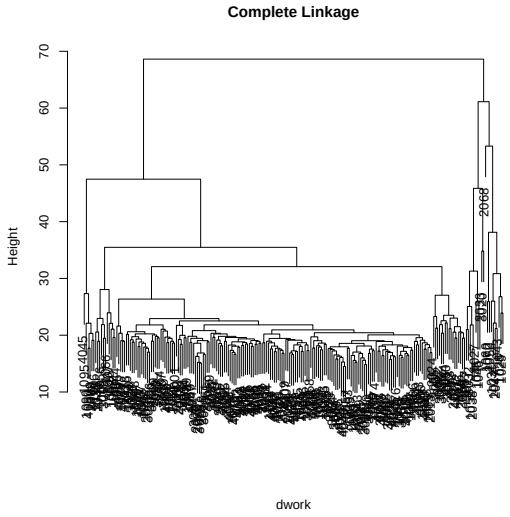


Figure 20: Complete Linkage Cluster Analysis For Mature Micro RNAs

3.5 Summary of Exploratory Analysis Results

- All sequence types (mature micro RNAs, stem loop pre micro RNAs and small nucleolar micro RNAs showed) some suggestion of exposure group related differences.
- These differences were mostly confined to exposure group 4, and represented a relatively small component of the total variation.
- The multivariate disposition of samples is fundamentally very similar for the three different types of micro RNA.

4 Differential Expression

4.1 Methods

Between group differences were evaluated with a linear model, and empirical Bayes moderated t tests[14] based on linear model parameter estimates. These were implemented using Smyth's limma methodology[18]. The following between group comparisons were performed:

1. Group 4 (Control) vs group 3 (Low exposure);
2. Group 4 (Control) vs group 2 (Moderate exposure);
3. Group 4 (Control) vs group 1 (High Exposure);

Individual p values for each micro RNA species were adjusted to maintain a false discovery rate of 5% using the Benjamini Hochberg procedure[20]. This procedure was adopted separately for each between groups comparison.

The empirical Bayes analysis was then repeated using health score as a covariate. Individual p values for the health score covariate for each micro RNA species were adjusted to maintain a false discovery rate of 5% using the Benjamini Hochberg procedure[20].

4.2 Mature Micro RNAs

Differentially expressed mature micro RNA species are shown in Table 5. One micro RNA appears to be differentially expressed in comparison between Group 1 and Group 4. There are 45 differentially expressed mature micro RNAs in comparison between Groups 1 (High exposure) and 4 (Control).

The largest fold change in expression is 4.7 fold. The mature micro RNAs re markedly more dynamic that the mRNA expression data. Even so, only seven species of micro RNAs achieved a greater than 2 fold change in expression in the High Exposure vs Control comparison.

Unfortunately, these 7 micro RNA species have strongly conserved high complementarity bindings with 4,403 mRNA species.

ID	logFC	AveExpr	t	P.Value	adj.P.Val	B	Contrast
hsa-miR-4282	-0.16	1.20	-3.16	0.00	0.05	-1.42	G1 v Control
hsa-miR-568	-0.17	1.18	-3.30	0.00	0.04	-1.02	G1 v Control
hsa-miR-519a-star	-0.18	1.02	-3.63	0.00	0.02	0.00	G1 v Control
hsa-miR-519c	-0.18	1.16	-3.56	0.00	0.02	-0.23	G1 v Control
hsa-miR-186	-0.19	1.25	-3.41	0.00	0.03	-0.68	G1 v Control
hsa-miR-376a	0.22	1.25	3.77	0.00	0.01	0.46	G1 v Control
hsa-miR-33b	-0.23	1.25	-3.27	0.00	0.04	-1.12	G1 v Control
hsa-miR-627	0.24	1.31	3.17	0.00	0.05	-1.42	G1 v Control
hsa-miR-4264	-0.25	1.27	-3.54	0.00	0.02	-0.29	G1 v Control
hsa-miR-25	0.30	12.36	3.27	0.00	0.04	-1.12	G1 v Control
hsa-miR-107	-0.33	11.58	-4.53	0.00	0.00	3.25	G1 v Control
hsa-miR-103	-0.34	11.93	-4.22	0.00	0.00	2.05	G1 v Control
hsa-miR-296	-0.35	1.81	-3.16	0.00	0.05	-1.44	G1 v Control
hsa-miR-208a	0.38	1.42	3.83	0.00	0.01	0.68	G1 v Control
hsa-miR-1468	0.42	1.53	3.41	0.00	0.03	-0.68	G1 v Control
hsa-miR-93	-0.46	10.64	-4.43	0.00	0.00	2.87	G1 v Control
hsa-miR-486	-0.50	3.21	-3.23	0.00	0.04	-1.23	G1 v Control
hsa-miR-766	-0.50	5.20	-3.43	0.00	0.03	-0.64	G1 v Control
hsa-miR-888	0.50	1.40	3.76	0.00	0.01	0.42	G1 v Control
hsa-miR-1234	-0.51	3.23	-3.68	0.00	0.01	0.17	G1 v Control
hsa-miR-1825	-0.52	4.73	-3.23	0.00	0.04	-1.22	G1 v Control
hsa-miR-3130-5p	-0.55	3.32	-3.14	0.00	0.05	-1.48	G1 v Control
hsa-miR-17	-0.58	10.16	-4.22	0.00	0.00	2.04	G1 v Control
hsa-miR-106a	-0.59	9.88	-4.06	0.00	0.01	1.46	G1 v Control
hsa-miR-18b	-0.61	2.30	-3.15	0.00	0.05	-1.48	G1 v Control
hsa-miR-1270	-0.62	3.77	-3.13	0.00	0.05	-1.52	G1 v Control
hsa-miR-548c	0.64	1.55	3.30	0.00	0.04	-1.04	G1 v Control
hsa-miR-559	0.70	2.10	3.67	0.00	0.01	0.15	G1 v Control
hsa-miR-27a	-0.75	3.03	-3.23	0.00	0.04	-1.25	G1 v Control
hsa-miR-129-3p	0.76	2.18	4.46	0.00	0.00	2.97	G1 v Control
hsa-miR-20b	-0.77	7.17	-3.97	0.00	0.01	1.15	G1 v Control
hsa-miR-1294	0.80	2.58	3.17	0.00	0.05	-1.39	G1 v Control
hsa-miR-1255a	0.85	1.90	3.96	0.00	0.01	1.13	G1 v Control
hsa-miR-20a	-0.92	7.67	-4.21	0.00	0.00	2.03	G1 v Control
hsa-miR-129-star	0.95	2.00	4.20	0.00	0.00	1.98	G1 v Control
hsa-miR-1255b	0.96	3.84	4.29	0.00	0.00	2.30	G1 v Control
hsa-miR-548f	0.96	1.76	3.62	0.00	0.02	-0.02	G1 v Control
hsa-miR-548e	0.99	1.70	3.70	0.00	0.01	0.23	G1 v Control
hsa-miR-548a-3p	1.05	3.35	3.26	0.00	0.04	-1.15	G1 v Control
hsa-miR-606	1.12	1.83	4.29	0.00	0.00	2.32	G1 v Control
hsa-miR-18a	-1.19	4.18	-4.45	0.00	0.00	2.94	G1 v Control
hsa-miR-576	1.54	2.03	4.58	0.00	0.00	3.45	G1 v Control
hsa-miR-335	1.63	3.12	3.95	0.00	0.01	1.10	G1 v Control
hsa-miR-3148	2.16	2.91	4.19	0.00	0.00	1.94	G1 v Control
hsa-miR-3128	2.22	3.11	4.43	0.00	0.00	2.84	G1 v Control
hsa-let-7d	-0.38	11.38	-3.77	0.00	0.05	0.48	Group 3 v Control
hsa-miR-93	-0.46	10.64	-3.76	0.00	0.05	0.46	Group 3 v Control
hsa-miR-17	-0.62	10.16	-3.82	0.00	0.05	0.64	Group 3 v Control

Table 5: Differentially Expressed Mature Micro RNA Species. The ID column is the micro RNA name as defined in the miRBase database[4] accessed via `http://www.mirbase.org/`.

ID	logFC	AveExpr	t	P.Value	adj.P.Val	B	Contrast
hsa-miR-106a	-0.66	9.88	-3.85	0.00	0.05	0.74	Group 3 v Control
hsa-miR-20a	-1.00	7.67	-3.89	0.00	0.05	0.89	Group 3 v Control

Table 5: Differentially Expressed Mature Micro RNA Species. The ID column is the micro RNA name as defined in the miRBase database[4] accessed via `http://www.mirbase.org/`.

Gene Ontology enrichment analyses for the target mRNAs are shown in Tables 6, 7, and 8 for Biological Process, Molecular Function and Cellular Component domains respectively. We note the for the Biological process domain there are a number of terms associated with RNA metabolic processes and transcription control, and also terms associated with central nervous system development and function. A primary ontology association with CNS development is not necessarily inconsistent with a role in CNS function and cognition in adulthood (this is certainly the case in the dentate gyrus which is known to have a high rate of neuro-genesis in adult rats).

The Molecular Function domain shows enrichment in several ontology terms associated with methylation / de-methylation and transcription factor binding. There are no obvious groupings in the Cellular Component domain.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0007399	nervous system development	0.00	1.46	484.96	602	1623
GO:2000112	regulation of cellular macromolecule biosynthetic process	0.00	1.37	775.70	914	2596
GO:0006996	organelle organization	0.00	1.40	645.42	774	2160
GO:0010556	regulation of macromolecule biosynthetic process	0.00	1.37	799.60	938	2676
GO:0010628	positive regulation of gene expression	0.00	1.57	302.09	395	1011
GO:0051252	regulation of RNA metabolic process	0.00	1.37	735.96	867	2463
GO:0009889	regulation of biosynthetic process	0.00	1.34	842.63	978	2820
GO:0031326	regulation of cellular biosynthetic process	0.00	1.34	834.26	968	2792
GO:0080090	regulation of primary metabolic process	0.00	1.31	1131.57	1276	3787
GO:0045893	positive regulation of transcription, DNA-templated	0.00	1.56	277.29	362	928
GO:0019219	regulation of nucleobase-containing compound metabolic process	0.00	1.33	884.16	1018	2959
GO:0044767	single-organism developmental process	0.00	1.30	1192.83	1338	3992
GO:0006351	transcription, DNA-templated	0.00	1.34	777.19	904	2601
GO:0051254	positive regulation of RNA metabolic process	0.00	1.53	292.83	378	980
GO:0060255	regulation of macromolecule metabolic process	0.00	1.30	1069.72	1208	3580
GO:0032502	developmental process	0.00	1.29	1204.78	1347	4032
GO:0031323	regulation of cellular metabolic process	0.00	1.29	1162.05	1303	3889
GO:0010558	negative regulation of macromolecule biosynthetic process	0.00	1.55	267.13	348	894
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	0.00	1.56	255.78	335	856
GO:0051171	regulation of nitrogen compound metabolic process	0.00	1.31	905.98	1036	3032
GO:0045935	positive regulation of nucleobase-containing compound metabolic process	0.00	1.49	326.59	414	1093
GO:2001141	regulation of RNA biosynthetic process	0.00	1.34	713.55	833	2388
GO:0010604	positive regulation of macromolecule metabolic process	0.00	1.39	518.43	624	1735
GO:0010468	regulation of gene expression	0.00	1.31	868.63	996	2907

Table 6: Biological Process Gene Ontology Enrichment for Mature Micro RNA Targets. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0006355	regulation of transcription, DNA-templated	0.00	1.34	704.28	822	2357
GO:1902680	positive regulation of RNA biosynthetic process	0.00	1.51	287.15	368	961
GO:0051173	positive regulation of nitrogen compound metabolic process	0.00	1.47	332.87	419	1114
GO:0048856	anatomical structure development	0.00	1.29	1031.47	1163	3452
GO:0010557	positive regulation of macromolecule biosynthetic process	0.00	1.47	324.50	409	1086
GO:0031327	negative regulation of cellular biosynthetic process	0.00	1.50	279.08	357	934
GO:0009890	negative regulation of biosynthetic process	0.00	1.50	282.07	360	944
GO:0009893	positive regulation of metabolic process	0.00	1.35	565.94	670	1894
GO:0032774	RNA biosynthetic process	0.00	1.31	795.42	913	2662
GO:1902589	single-organism organelle organization	0.00	1.40	428.78	521	1435
GO:0010629	negative regulation of gene expression	0.00	1.51	253.09	325	847
GO:0009891	positive regulation of biosynthetic process	0.00	1.42	350.80	433	1174
GO:0031328	positive regulation of cellular biosynthetic process	0.00	1.43	344.52	426	1153
GO:0050794	regulation of cellular process	0.00	1.25	1969.42	2101	6591
GO:0034654	nucleobase-containing compound biosynthetic process	0.00	1.28	865.34	981	2896
GO:0021537	telencephalon development	0.00	2.44	46.02	78	154
GO:0010605	negative regulation of macromolecule metabolic process	0.00	1.40	388.15	473	1299
GO:0031325	positive regulation of cellular metabolic process	0.00	1.34	537.25	634	1798
GO:0006366	transcription from RNA polymerase II promoter	0.00	1.39	400.10	485	1339
GO:0019222	regulation of metabolic process	0.00	1.25	1284.56	1410	4299
GO:0009653	anatomical structure morphogenesis	0.00	1.32	569.22	666	1905
GO:0032989	cellular component morphogenesis	0.00	1.45	289.54	362	969
GO:0007275	multicellular organismal development	0.00	1.25	1060.76	1179	3550
GO:0030182	neuron differentiation	0.00	1.44	290.74	363	973
GO:0044260	cellular macromolecule metabolic process	0.00	1.23	1640.44	1767	5490
GO:0044271	cellular nitrogen compound biosynthetic process	0.00	1.27	896.12	1008	2999

Table 6: Biological Process Gene Ontology Enrichment for Mature Micro RNA Targets. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0019438	aromatic compound biosynthetic process	0.00	1.27	881.77	993	2951
GO:1902679	negative regulation of RNA biosynthetic process	0.00	1.50	230.38	295	771
GO:0016043	cellular component organization	0.00	1.24	1188.65	1309	3978
GO:0006357	regulation of transcription from RNA polymerase II promoter	0.00	1.40	353.19	431	1182
GO:0045892	negative regulation of transcription, DNA-templated	0.00	1.50	225.60	289	755
GO:0006464	cellular protein modification process	0.00	1.29	681.87	782	2282
GO:0036211	protein modification process	0.00	1.29	681.87	782	2282
GO:0021543	pallium development	0.00	2.84	29.58	54	99
GO:0048699	generation of neurons	0.00	1.41	314.34	387	1052
GO:0051253	negative regulation of RNA metabolic process	0.00	1.48	237.85	302	796
GO:0009059	macromolecule biosynthetic process	0.00	1.25	1041.93	1156	3487
GO:0007010	cytoskeleton organization	0.00	1.49	220.22	282	737
GO:0034645	cellular macromolecule biosynthetic process	0.00	1.25	1012.65	1125	3389
GO:0000902	cell morphogenesis	0.00	1.44	275.20	343	921
GO:0018130	heterocycle biosynthetic process	0.00	1.26	880.58	988	2947
GO:0048522	positive regulation of cellular process	0.00	1.26	883.57	991	2957
GO:0048731	system development	0.00	1.25	884.46	991	2960
GO:0065007	biological regulation	0.00	1.23	2185.16	2301	7313
GO:0000904	cell morphogenesis involved in differentiation	0.00	1.50	201.99	260	676
GO:0016070	RNA metabolic process	0.00	1.25	948.11	1056	3173
GO:0045934	negative regulation of nucleobase-containing compound metabolic process	0.00	1.44	261.16	326	874
GO:0016568	chromatin modification	0.00	1.65	129.08	176	432
GO:0071840	cellular component organization or biogenesis	0.00	1.23	1207.47	1322	4041
GO:0006139	nucleobase-containing compound metabolic process	0.00	1.22	1294.12	1410	4331
GO:0045944	positive regulation of transcription from RNA polymerase II promoter	0.00	1.50	195.72	252	655

Table 6: Biological Process Gene Ontology Enrichment for Mature Micro RNA Targets. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

ID	logFC	AveExpr	t	P.Value	adj.P.Val	B	Contrast
hsa-miR-4282	-0.16	1.20	-3.16	0.00	0.05	-1.42	G1 v Control
hsa-miR-568	-0.17	1.18	-3.30	0.00	0.04	-1.02	G1 v Control
hsa-miR-519a-star	-0.18	1.02	-3.63	0.00	0.02	0.00	G1 v Control
hsa-miR-519c	-0.18	1.16	-3.56	0.00	0.02	-0.23	G1 v Control
hsa-miR-186	-0.19	1.25	-3.41	0.00	0.03	-0.68	G1 v Control
hsa-miR-376a	0.22	1.25	3.77	0.00	0.01	0.46	G1 v Control
hsa-miR-33b	-0.23	1.25	-3.27	0.00	0.04	-1.12	G1 v Control
hsa-miR-627	0.24	1.31	3.17	0.00	0.05	-1.42	G1 v Control
hsa-miR-4264	-0.25	1.27	-3.54	0.00	0.02	-0.29	G1 v Control
hsa-miR-25	0.30	12.36	3.27	0.00	0.04	-1.12	G1 v Control
hsa-miR-107	-0.33	11.58	-4.53	0.00	0.00	3.25	G1 v Control
hsa-miR-103	-0.34	11.93	-4.22	0.00	0.00	2.05	G1 v Control
hsa-miR-296	-0.35	1.81	-3.16	0.00	0.05	-1.44	G1 v Control
hsa-miR-208a	0.38	1.42	3.83	0.00	0.01	0.68	G1 v Control
hsa-miR-1468	0.42	1.53	3.41	0.00	0.03	-0.68	G1 v Control
hsa-miR-93	-0.46	10.64	-4.43	0.00	0.00	2.87	G1 v Control
hsa-miR-486	-0.50	3.21	-3.23	0.00	0.04	-1.23	G1 v Control
hsa-miR-766	-0.50	5.20	-3.43	0.00	0.03	-0.64	G1 v Control
hsa-miR-888	0.50	1.40	3.76	0.00	0.01	0.42	G1 v Control
hsa-miR-1234	-0.51	3.23	-3.68	0.00	0.01	0.17	G1 v Control
hsa-miR-1825	-0.52	4.73	-3.23	0.00	0.04	-1.22	G1 v Control
hsa-miR-3130-5p	-0.55	3.32	-3.14	0.00	0.05	-1.48	G1 v Control
hsa-miR-17	-0.58	10.16	-4.22	0.00	0.00	2.04	G1 v Control
hsa-miR-106a	-0.59	9.88	-4.06	0.00	0.01	1.46	G1 v Control
hsa-miR-18b	-0.61	2.30	-3.15	0.00	0.05	-1.48	G1 v Control
hsa-miR-1270	-0.62	3.77	-3.13	0.00	0.05	-1.52	G1 v Control
hsa-miR-548c	0.64	1.55	3.30	0.00	0.04	-1.04	G1 v Control
hsa-miR-559	0.70	2.10	3.67	0.00	0.01	0.15	G1 v Control
hsa-miR-27a	-0.75	3.03	-3.23	0.00	0.04	-1.25	G1 v Control
hsa-miR-129-3p	0.76	2.18	4.46	0.00	0.00	2.97	G1 v Control
hsa-miR-20b	-0.77	7.17	-3.97	0.00	0.01	1.15	G1 v Control
hsa-miR-1294	0.80	2.58	3.17	0.00	0.05	-1.39	G1 v Control
hsa-miR-1255a	0.85	1.90	3.96	0.00	0.01	1.13	G1 v Control
hsa-miR-20a	-0.92	7.67	-4.21	0.00	0.00	2.03	G1 v Control
hsa-miR-129-star	0.95	2.00	4.20	0.00	0.00	1.98	G1 v Control
hsa-miR-1255b	0.96	3.84	4.29	0.00	0.00	2.30	G1 v Control
hsa-miR-548f	0.96	1.76	3.62	0.00	0.02	-0.02	G1 v Control
hsa-miR-548e	0.99	1.70	3.70	0.00	0.01	0.23	G1 v Control
hsa-miR-548a-3p	1.05	3.35	3.26	0.00	0.04	-1.15	G1 v Control
hsa-miR-606	1.12	1.83	4.29	0.00	0.00	2.32	G1 v Control
hsa-miR-18a	-1.19	4.18	-4.45	0.00	0.00	2.94	G1 v Control
hsa-miR-576	1.54	2.03	4.58	0.00	0.00	3.45	G1 v Control
hsa-miR-335	1.63	3.12	3.95	0.00	0.01	1.10	G1 v Control
hsa-miR-3148	2.16	2.91	4.19	0.00	0.00	1.94	G1 v Control
hsa-miR-3128	2.22	3.11	4.43	0.00	0.00	2.84	G1 v Control
hsa-let-7d	-0.38	11.38	-3.77	0.00	0.05	0.48	Group 3 v Control
hsa-miR-93	-0.46	10.64	-3.76	0.00	0.05	0.46	Group 3 v Control
hsa-miR-17	-0.62	10.16	-3.82	0.00	0.05	0.64	Group 3 v Control

Table 5: Differentially Expressed Mature Micro RNA Species. The ID column is the micro RNA name as defined in the miRBase database[4] accessed via `http://www.mirbase.org/`.

Unfortunately, these 7 micro RNA species have strongly conserved high complementarity bindings with 4,403 mRNA species.

3.5 Summary of Exploratory Analysis Results

- All sequence types (mature micro RNAs, stem loop pre micro RNAs and small nucleolar micro RNAs showed) some suggestion of exposure group related differences.
- These differences were mostly confined to exposure group 4, and represented a relatively small component of the total variation.
- The multivariate disposition of samples is fundamentally very similar for the three different types of micro RNA.

4 Differential Expression

4.1 Methods

Between group differences were evaluated with a linear model, and empirical Bayes moderated t tests[14] based on linear model parameter estimates. These were implemented using Smyth's limma methodology[18]. The following between group comparisons were performed:

1. Group 4 (Control) vs group 3 (Low exposure);
2. Group 4 (Control) vs group 2 (Moderate exposure);
3. Group 4 (Control) vs group 1 (High Exposure);

Individual p values for each micro RNA species were adjusted to maintain a false discovery rate of 5% using the Benjamini Hochberg procedure[20]. This procedure was adopted separately for each between groups comparison.

The empirical Bayes analysis was then repeated using health score as a covariate. Individual p values for the health score covariate for each micro RNA species were adjusted to maintain a false discovery rate of 5% using the Benjamini Hochberg procedure[20].

4.2 Mature Micro RNAs

Differentially expressed mature micro RNA species are shown in Table 5. One micro RNA appears to be differentially expressed in comparison between Group 1 and Group 4. There are 45 differentially expressed mature micro RNAs in comparison between Groups 1 (High exposure) and 4 (Control).

The largest fold change in expression is 4.7 fold. The mature micro RNAs re markedly more dynamic that the mRNA expression data. Even so, only seven species of micro RNAs achieved a greater than 2 fold change in expression in the High Exposure vs Control comparison.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0007173	epidermal growth factor receptor signaling pathway	0.00	1.69	58.86	82	197
GO:1901699	cellular response to nitrogen compound	0.00	1.45	119.82	152	401
GO:0038127	ERBB signaling pathway	0.00	1.68	59.76	83	200
GO:0009058	biosynthetic process	0.00	1.16	1261.85	1344	4223
GO:1901701	cellular response to oxygen-containing compound	0.00	1.35	182.57	221	611
GO:0001568	blood vessel development	0.00	1.46	113.25	144	379
GO:0007411	axon guidance	0.00	1.49	100.10	129	335
GO:0097485	neuron projection guidance	0.00	1.49	100.10	129	335
GO:0035335	peptidyl-tyrosine dephosphorylation	0.00	2.53	17.33	30	58
GO:0001944	vasculature development	0.00	1.39	145.82	180	488
GO:0001958	endochondral ossification	0.00	4.19	7.47	16	25
GO:0036075	replacement ossification	0.00	4.19	7.47	16	25
GO:0045595	regulation of cell differentiation	0.00	1.27	305.38	353	1022
GO:0070848	response to growth factor	0.00	1.36	171.51	208	574
GO:0033043	regulation of organelle organization	0.00	1.36	171.51	208	574
GO:0044707	single-multicellular organism process	0.00	1.15	1343.13	1423	4495
GO:0001654	eye development	0.00	1.57	73.51	98	246
GO:0016570	histone modification	0.00	1.53	83.07	109	278
GO:0006807	nitrogen compound metabolic process	0.00	1.15	1463.25	1544	4897
GO:0048864	stem cell development	0.00	1.71	50.50	71	169
GO:0019827	stem cell maintenance	0.00	2.04	28.39	44	95
GO:0043254	regulation of protein complex assembly	0.00	1.64	59.16	81	198
GO:0007015	actin filament organization	0.00	1.60	66.04	89	221
GO:0016458	gene silencing	0.00	2.03	27.79	43	93
GO:0032501	multicellular organismal process	0.00	1.14	1388.55	1467	4647
GO:0007612	learning	0.00	2.06	26.29	41	88

Table 6: Biological Process Gene Ontology Enrichment for Mature Micro RNA Targets. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0034968	histone lysine methylation	0.00	2.44	17.03	29	57
GO:0001709	cell fate determination	0.00	3.14	10.46	20	35
GO:0006208	pyrimidine nucleobase catabolic process	0.00	Inf	1.79	6	6
GO:0071310	cellular response to organic substance	0.00	1.22	409.36	461	1370
GO:2000027	regulation of organ morphogenesis	0.00	1.76	42.73	61	143
GO:0060284	regulation of cell development	0.00	1.35	157.47	191	527
GO:0007167	enzyme linked receptor protein signaling pathway	0.00	1.28	251.59	293	842
GO:0043525	positive regulation of neuron apoptotic process	0.00	2.88	11.95	22	40
GO:2000826	regulation of heart morphogenesis	0.00	3.77	7.77	16	26
GO:0071363	cellular response to growth factor stimulus	0.00	1.34	166.73	201	558
GO:0006900	membrane budding	0.00	2.69	13.45	24	45
GO:0048592	eye morphogenesis	0.00	1.85	34.66	51	116
GO:0050767	regulation of neurogenesis	0.00	1.40	125.20	155	419
GO:0035051	cardiocyte differentiation	0.00	1.96	28.98	44	97
GO:0071417	cellular response to organonitrogen compound	0.00	1.42	110.86	139	371
GO:0071229	cellular response to acid	0.00	1.83	35.56	52	119
GO:0048514	blood vessel morphogenesis	0.00	1.45	95.92	122	321
GO:0006470	protein dephosphorylation	0.00	1.77	39.74	57	133
GO:0030534	adult behavior	0.00	1.87	33.17	49	111
GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	0.00	1.32	181.08	216	606

Table 6: Biological Process Gene Ontology Enrichment for Mature Micro RNA Targets. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0007399	nervous system development	0.00	1.46	484.96	602	1623
GO:2000112	regulation of cellular macromolecule biosynthetic process	0.00	1.37	775.70	914	2596
GO:0006996	organelle organization	0.00	1.40	645.42	774	2160
GO:0010556	regulation of macromolecule biosynthetic process	0.00	1.37	799.60	938	2676
GO:0010628	positive regulation of gene expression	0.00	1.57	302.09	395	1011
GO:0051252	regulation of RNA metabolic process	0.00	1.37	735.96	867	2463
GO:0009889	regulation of biosynthetic process	0.00	1.34	842.63	978	2820
GO:0031326	regulation of cellular biosynthetic process	0.00	1.34	834.26	968	2792
GO:0080090	regulation of primary metabolic process	0.00	1.31	1131.57	1276	3787
GO:0045893	positive regulation of transcription, DNA-templated	0.00	1.56	277.29	362	928
GO:0019219	regulation of nucleobase-containing compound metabolic process	0.00	1.33	884.16	1018	2959
GO:0044767	single-organism developmental process	0.00	1.30	1192.83	1338	3992
GO:0006351	transcription, DNA-templated	0.00	1.34	777.19	904	2601
GO:0051254	positive regulation of RNA metabolic process	0.00	1.53	292.83	378	980
GO:0060255	regulation of macromolecule metabolic process	0.00	1.30	1069.72	1208	3580
GO:0032502	developmental process	0.00	1.29	1204.78	1347	4032
GO:0031323	regulation of cellular metabolic process	0.00	1.29	1162.05	1303	3889
GO:0010558	negative regulation of macromolecule biosynthetic process	0.00	1.55	267.13	348	894
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	0.00	1.56	255.78	335	856
GO:0051171	regulation of nitrogen compound metabolic process	0.00	1.31	905.98	1036	3032
GO:0045935	positive regulation of nucleobase-containing compound metabolic process	0.00	1.49	326.59	414	1093
GO:2001141	regulation of RNA biosynthetic process	0.00	1.34	713.55	833	2388
GO:0010604	positive regulation of macromolecule metabolic process	0.00	1.39	518.43	624	1735
GO:0010468	regulation of gene expression	0.00	1.31	868.63	996	2907

Table 7: Molecular Function Gene Ontology Enrichment for Mature Micro RNA Targets. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0006355	regulation of transcription, DNA-templated	0.00	1.34	704.28	822	2357
GO:1902680	positive regulation of RNA biosynthetic process	0.00	1.51	287.15	368	961
GO:0051173	positive regulation of nitrogen compound metabolic process	0.00	1.47	332.87	419	1114
GO:0048856	anatomical structure development	0.00	1.29	1031.47	1163	3452
GO:0010557	positive regulation of macromolecule biosynthetic process	0.00	1.47	324.50	409	1086
GO:0031327	negative regulation of cellular biosynthetic process	0.00	1.50	279.08	357	934
GO:0009890	negative regulation of biosynthetic process	0.00	1.50	282.07	360	944
GO:0009893	positive regulation of metabolic process	0.00	1.35	565.94	670	1894
GO:0032774	RNA biosynthetic process	0.00	1.31	795.42	913	2662
GO:1902589	single-organism organelle organization	0.00	1.40	428.78	521	1435
GO:0010629	negative regulation of gene expression	0.00	1.51	253.09	325	847
GO:0009891	positive regulation of biosynthetic process	0.00	1.42	350.80	433	1174
GO:0031328	positive regulation of cellular biosynthetic process	0.00	1.43	344.52	426	1153
GO:0050794	regulation of cellular process	0.00	1.25	1969.42	2101	6591
GO:0034654	nucleobase-containing compound biosynthetic process	0.00	1.28	865.34	981	2896
GO:0021537	telencephalon development	0.00	2.44	46.02	78	154
GO:0010605	negative regulation of macromolecule metabolic process	0.00	1.40	388.15	473	1299
GO:0031325	positive regulation of cellular metabolic process	0.00	1.34	537.25	634	1798
GO:0006366	transcription from RNA polymerase II promoter	0.00	1.39	400.10	485	1339
GO:0019222	regulation of metabolic process	0.00	1.25	1284.56	1410	4299
GO:0009653	anatomical structure morphogenesis	0.00	1.32	569.22	666	1905
GO:0032989	cellular component morphogenesis	0.00	1.45	289.54	362	969
GO:0007275	multicellular organismal development	0.00	1.25	1060.76	1179	3550
GO:0030182	neuron differentiation	0.00	1.44	290.74	363	973
GO:0044260	cellular macromolecule metabolic process	0.00	1.23	1640.44	1767	5490
GO:0044271	cellular nitrogen compound biosynthetic process	0.00	1.27	896.12	1008	2999

Table 7: Molecular Function Gene Ontology Enrichment for Mature Micro RNA Targets. The ‘GO ID’ is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0019438	aromatic compound biosynthetic process	0.00	1.27	881.77	993	2951
GO:1902679	negative regulation of RNA biosynthetic process	0.00	1.50	230.38	295	771
GO:0016043	cellular component organization	0.00	1.24	1188.65	1309	3978
GO:0006357	regulation of transcription from RNA polymerase II promoter	0.00	1.40	353.19	431	1182
GO:0045892	negative regulation of transcription, DNA-templated	0.00	1.50	225.60	289	755
GO:0006464	cellular protein modification process	0.00	1.29	681.87	782	2282
GO:0036211	protein modification process	0.00	1.29	681.87	782	2282
GO:0021543	pallium development	0.00	2.84	29.58	54	99
GO:0048699	generation of neurons	0.00	1.41	314.34	387	1052
GO:0051253	negative regulation of RNA metabolic process	0.00	1.48	237.85	302	796
GO:0009059	macromolecule biosynthetic process	0.00	1.25	1041.93	1156	3487
GO:0007010	cytoskeleton organization	0.00	1.49	220.22	282	737
GO:0034645	cellular macromolecule biosynthetic process	0.00	1.25	1012.65	1125	3389
GO:0000902	cell morphogenesis	0.00	1.44	275.20	343	921
GO:0018130	heterocycle biosynthetic process	0.00	1.26	880.58	988	2947
GO:0048522	positive regulation of cellular process	0.00	1.26	883.57	991	2957
GO:0048731	system development	0.00	1.25	884.46	991	2960
GO:0065007	biological regulation	0.00	1.23	2185.16	2301	7313
GO:0000904	cell morphogenesis involved in differentiation	0.00	1.50	201.99	260	676
GO:0016070	RNA metabolic process	0.00	1.25	948.11	1056	3173
GO:0045934	negative regulation of nucleobase-containing compound metabolic process	0.00	1.44	261.16	326	874
GO:0016568	chromatin modification	0.00	1.65	129.08	176	432
GO:0071840	cellular component organization or biogenesis	0.00	1.23	1207.47	1322	4041
GO:0006139	nucleobase-containing compound metabolic process	0.00	1.22	1294.12	1410	4331
GO:0045944	positive regulation of transcription from RNA polymerase II promoter	0.00	1.50	195.72	252	655

Table 7: Molecular Function Gene Ontology Enrichment for Mature Micro RNA Targets. The ‘GO ID’ is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0022008	neurogenesis	0.00	1.38	331.97	403	1111
GO:1901362	organic cyclic compound biosynthetic process	0.00	1.24	904.48	1009	3027
GO:0030900	forebrain development	0.00	1.78	90.84	130	304
GO:0050789	regulation of biological process	0.00	1.22	2068.03	2183	6921
GO:0051172	negative regulation of nitrogen compound metabolic process	0.00	1.42	265.94	330	890
GO:0006325	chromatin organization	0.00	1.58	146.41	195	490
GO:0007507	heart development	0.00	1.69	107.57	149	360
GO:0048667	cell morphogenesis involved in neuron differentiation	0.00	1.60	134.16	180	449
GO:0048468	cell development	0.00	1.32	432.07	509	1446
GO:0009892	negative regulation of metabolic process	0.00	1.33	421.91	498	1412
GO:0043412	macromolecule modification	0.00	1.26	704.28	797	2357
GO:0007417	central nervous system development	0.00	1.47	199.60	254	668
GO:0048523	negative regulation of cellular process	0.00	1.25	798.41	895	2672
GO:0048869	cellular developmental process	0.00	1.25	793.63	890	2656
GO:0031324	negative regulation of cellular metabolic process	0.00	1.33	389.64	462	1304
GO:0090304	nucleic acid metabolic process	0.00	1.22	1069.12	1172	3578
GO:0048518	positive regulation of biological process	0.00	1.22	983.67	1084	3292
GO:0044708	single-organism behavior	0.00	1.76	83.67	119	280
GO:0051276	chromosome organization	0.00	1.47	193.03	245	646
GO:0031175	neuron projection development	0.00	1.45	205.58	259	688
GO:0030030	cell projection organization	0.00	1.38	277.59	338	929
GO:0010467	gene expression	0.00	1.20	1237.95	1342	4143
GO:0044087	regulation of cellular component biogenesis	0.00	1.61	112.65	152	377
GO:0048812	neuron projection morphogenesis	0.00	1.49	163.45	210	547
GO:0030154	cell differentiation	0.00	1.23	747.61	836	2502
GO:0061564	axon development	0.00	1.55	132.07	174	442

Table 7: Molecular Function Gene Ontology Enrichment for Mature Micro RNA Targets. The ‘GO ID’ is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0021987	cerebral cortex development	0.00	2.88	21.22	39	71
GO:0048519	negative regulation of biological process	0.00	1.22	866.53	958	2900
GO:0050890	cognition	0.00	2.00	49.00	75	164
GO:0030036	actin cytoskeleton organization	0.00	1.57	119.82	159	401
GO:0006417	regulation of translation	0.00	1.84	63.05	92	211
GO:0048666	neuron development	0.00	1.43	196.31	245	657
GO:0072358	cardiovascular system development	0.00	1.43	191.83	240	642
GO:0072359	circulatory system development	0.00	1.43	191.83	240	642
GO:0032990	cell part morphogenesis	0.00	1.41	199.90	248	669
GO:0006725	cellular aromatic compound metabolic process	0.00	1.18	1332.37	1430	4459
GO:0051128	regulation of cellular component organization	0.00	1.29	388.45	452	1300
GO:0010608	posttranscriptional regulation of gene expression	0.00	1.58	105.78	141	354
GO:0007049	cell cycle	0.00	1.29	386.95	450	1295
GO:0034641	cellular nitrogen compound metabolic process	0.00	1.18	1377.49	1474	4610
GO:0046483	heterocycle metabolic process	0.00	1.18	1328.49	1424	4446
GO:0051493	regulation of cytoskeleton organization	0.00	1.65	81.57	112	273
GO:0048858	cell projection morphogenesis	0.00	1.40	197.21	243	660
GO:0007420	brain development	0.00	1.45	152.39	193	510
GO:0044267	cellular protein metabolic process	0.00	1.20	859.06	942	2875
GO:0017148	negative regulation of translation	0.00	2.67	19.72	35	66
GO:0007611	learning or memory	0.00	1.92	45.42	68	152
GO:0043170	macromolecule metabolic process	0.00	1.17	1797.91	1892	6017
GO:0021542	dentate gyrus development	0.00	12.94	3.88	11	13
GO:0044249	cellular biosynthetic process	0.00	1.18	1223.61	1314	4095
GO:0032956	regulation of actin cytoskeleton organization	0.00	1.80	56.18	81	188
GO:0050793	regulation of developmental process	0.00	1.26	420.12	482	1406

Table 7: Molecular Function Gene Ontology Enrichment for Mature Micro RNA Targets. The ‘GO ID’ is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0035019	somatic stem cell maintenance	0.00	3.62	11.35	23	38
GO:0009887	organ morphogenesis	0.00	1.36	213.35	259	714
GO:0042481	regulation of odontogenesis	0.00	5.89	6.27	15	21
GO:0007409	axonogenesis	0.00	1.44	143.43	181	480
GO:0042692	muscle cell differentiation	0.00	1.64	76.20	104	255
GO:0030029	actin filament-based process	0.00	1.46	132.07	168	442
GO:0016569	covalent chromatin modification	0.00	1.59	84.86	114	284
GO:0007423	sensory organ development	0.00	1.49	118.03	152	395
GO:0071230	cellular response to amino acid stimulus	0.00	3.33	12.25	24	41
GO:1901360	organic cyclic compound metabolic process	0.00	1.16	1373.31	1462	4596
GO:0000122	negative regulation of transcription from RNA polymerase II promoter	0.00	1.43	144.02	181	482
GO:0042487	regulation of odontogenesis of dentin-containing tooth	0.00	21.17	2.99	9	10
GO:0009790	embryo development	0.00	1.32	247.11	294	827
GO:1901576	organic substance biosynthetic process	0.00	1.16	1245.72	1331	4169
GO:0006468	protein phosphorylation	0.00	1.28	309.86	361	1037
GO:0071495	cellular response to endogenous stimulus	0.00	1.33	225.00	269	753
GO:0045165	cell fate commitment	0.00	1.68	61.85	86	207
GO:0022402	cell cycle process	0.00	1.29	290.74	340	973
GO:0048332	mesoderm morphogenesis	0.00	2.74	16.14	29	54
GO:0021953	central nervous system neuron differentiation	0.00	1.86	42.73	63	143
GO:0032970	regulation of actin filament-based process	0.00	1.65	67.83	93	227
GO:0048568	embryonic organ development	0.00	1.51	102.49	133	343
GO:0032268	regulation of cellular protein metabolic process	0.00	1.26	368.73	423	1234
GO:0061061	muscle structure development	0.00	1.45	124.00	157	415
GO:0003007	heart morphogenesis	0.00	1.77	49.60	71	166
GO:0022607	cellular component assembly	0.00	1.23	434.46	492	1454

Table 7: Molecular Function Gene Ontology Enrichment for Mature Micro RNA Targets. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0007173	epidermal growth factor receptor signaling pathway	0.00	1.69	58.86	82	197
GO:1901699	cellular response to nitrogen compound	0.00	1.45	119.82	152	401
GO:0038127	ERBB signaling pathway	0.00	1.68	59.76	83	200
GO:0009058	biosynthetic process	0.00	1.16	1261.85	1344	4223
GO:1901701	cellular response to oxygen-containing compound	0.00	1.35	182.57	221	611
GO:0001568	blood vessel development	0.00	1.46	113.25	144	379
GO:0007411	axon guidance	0.00	1.49	100.10	129	335
GO:0097485	neuron projection guidance	0.00	1.49	100.10	129	335
GO:0035335	peptidyl-tyrosine dephosphorylation	0.00	2.53	17.33	30	58
GO:0001944	vasculature development	0.00	1.39	145.82	180	488
GO:0001958	endochondral ossification	0.00	4.19	7.47	16	25
GO:0036075	replacement ossification	0.00	4.19	7.47	16	25
GO:0045595	regulation of cell differentiation	0.00	1.27	305.38	353	1022
GO:0070848	response to growth factor	0.00	1.36	171.51	208	574
GO:0033043	regulation of organelle organization	0.00	1.36	171.51	208	574
GO:0044707	single-multicellular organism process	0.00	1.15	1343.13	1423	4495
GO:0001654	eye development	0.00	1.57	73.51	98	246
GO:0016570	histone modification	0.00	1.53	83.07	109	278
GO:0006807	nitrogen compound metabolic process	0.00	1.15	1463.25	1544	4897
GO:0048864	stem cell development	0.00	1.71	50.50	71	169
GO:0019827	stem cell maintenance	0.00	2.04	28.39	44	95
GO:0043254	regulation of protein complex assembly	0.00	1.64	59.16	81	198
GO:0007015	actin filament organization	0.00	1.60	66.04	89	221
GO:0016458	gene silencing	0.00	2.03	27.79	43	93
GO:0032501	multicellular organismal process	0.00	1.14	1388.55	1467	4647
GO:0007612	learning	0.00	2.06	26.29	41	88

Table 7: Molecular Function Gene Ontology Enrichment for Mature Micro RNA Targets. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0034968	histone lysine methylation	0.00	2.44	17.03	29	57
GO:0001709	cell fate determination	0.00	3.14	10.46	20	35
GO:0006208	pyrimidine nucleobase catabolic process	0.00	Inf	1.79	6	6
GO:0071310	cellular response to organic substance	0.00	1.22	409.36	461	1370
GO:2000027	regulation of organ morphogenesis	0.00	1.76	42.73	61	143
GO:0060284	regulation of cell development	0.00	1.35	157.47	191	527
GO:0007167	enzyme linked receptor protein signaling pathway	0.00	1.28	251.59	293	842
GO:0043525	positive regulation of neuron apoptotic process	0.00	2.88	11.95	22	40
GO:2000826	regulation of heart morphogenesis	0.00	3.77	7.77	16	26
GO:0071363	cellular response to growth factor stimulus	0.00	1.34	166.73	201	558
GO:0006900	membrane budding	0.00	2.69	13.45	24	45
GO:0048592	eye morphogenesis	0.00	1.85	34.66	51	116
GO:0050767	regulation of neurogenesis	0.00	1.40	125.20	155	419
GO:0035051	cardiocyte differentiation	0.00	1.96	28.98	44	97
GO:0071417	cellular response to organonitrogen compound	0.00	1.42	110.86	139	371
GO:0071229	cellular response to acid	0.00	1.83	35.56	52	119
GO:0048514	blood vessel morphogenesis	0.00	1.45	95.92	122	321
GO:0006470	protein dephosphorylation	0.00	1.77	39.74	57	133
GO:0030534	adult behavior	0.00	1.87	33.17	49	111
GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	0.00	1.32	181.08	216	606

Table 7: Molecular Function Gene Ontology Enrichment for Mature Micro RNA Targets. The ‘GO ID’ is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0005634	nucleus	0.00	1.24	1435.67	1573	4851
GO:0005856	cytoskeleton	0.00	1.32	443.63	523	1499
GO:0005622	intracellular	0.00	1.25	2961.61	3060	10007
GO:0044424	intracellular part	0.00	1.24	2936.16	3032	9921
GO:0043229	intracellular organelle	0.00	1.20	2546.98	2650	8606
GO:0005654	nucleoplasm	0.00	1.31	373.49	440	1262
GO:0044428	nuclear part	0.00	1.24	625.05	706	2112
GO:0031981	nuclear lumen	0.00	1.26	528.87	604	1787
GO:0043226	organelle	0.00	1.19	2681.04	2776	9059
GO:0044464	cell part	0.00	1.28	3330.96	3397	11255
GO:0005623	cell	0.00	1.28	3331.25	3397	11256
GO:0043228	non-membrane-bounded organelle	0.00	1.21	780.73	864	2638
GO:0043232	intracellular non-membrane-bounded organelle	0.00	1.21	780.73	864	2638
GO:0042995	cell projection	0.00	1.29	348.63	408	1178
GO:0031519	PcG protein complex	0.00	3.58	11.84	24	40
GO:0005815	microtubule organizing center	0.00	1.49	125.48	162	424
GO:0055037	recycling endosome	0.00	2.61	20.42	36	69
GO:0030054	cell junction	0.00	1.37	208.06	254	703
GO:0015630	microtubule cytoskeleton	0.00	1.35	225.52	272	762
GO:0031252	cell leading edge	0.00	1.61	79.61	108	269
GO:0043231	intracellular membrane-bounded organelle	0.00	1.15	2308.14	2396	7799
GO:0043005	neuron projection	0.00	1.36	184.38	224	623
GO:0030027	lamellipodium	0.00	1.92	37.29	56	126
GO:0005720	nuclear heterochromatin	0.00	4.47	6.81	15	23
GO:0005912	adherens junction	0.00	1.69	54.46	76	184

Table 8: Cellular Component Gene Ontology Enrichment for Mature Micro RNA Targets. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0043227	membrane-bounded organelle	0.00	1.14	2498.74	2578	8443
GO:0016342	catenin complex	0.00	Inf	1.78	6	6
GO:0005813	centrosome	0.00	1.46	95.59	122	323
GO:0055038	recycling endosome membrane	0.00	3.30	9.17	18	31

Table 8: Cellular Component Gene Ontology Enrichment for Mature Micro RNA Targets. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

4.3 Stem Loop pre Micro RNAs

Differentially expressed stem-loop pre micro RNAs are shown in Table 9. There are several micro RNA species in common with the mature micro RNAs. These pre micro RNAs are associated with 4,715 target mRNAs.

Gene Ontology enrichment analyses for the target mRNAs are shown in Tables 10, 11, and 12 for Biological Process, Molecular Function and Cellular Component domains respectively. We note the for the Biological process domain there are a number of terms associated with RNA metabolic processes and transcription control, and also a few terms associated with central nervous system development and function. A primary ontology association with CNS development is not necessarily inconsistent with a role in CNS function and cognition in adulthood.

The Molecular Function domain shows enrichment for several terms associated (either as child or ancestor nodes) with nucleic acid binding. The Cellular Component domain shows enrichment for cell nucleus terms.

ID	logFC	AveExpr	t	P.Value	adj.P.Val	B	Contrast
hsa-miR-1297	-0.17	1.34	-3.39	0.00	0.03	-0.73	G1 v Control
hsa-miR-3115	-0.17	1.14	-3.43	0.00	0.03	-0.63	G1 v Control
hsa-miR-100	-0.18	1.21	-3.53	0.00	0.03	-0.29	G1 v Control
hsa-miR-513a-1	-0.18	1.31	-3.35	0.00	0.03	-0.86	G1 v Control
hsa-miR-4266	-0.19	1.40	-3.23	0.00	0.04	-1.21	G1 v Control
hsa-miR-628	0.20	1.28	3.22	0.00	0.04	-1.24	G1 v Control
hsa-miR-1-1	-0.21	1.31	-3.64	0.00	0.02	0.04	G1 v Control
hsa-miR-520e	0.22	1.29	3.44	0.00	0.03	-0.58	G1 v Control
hsa-miR-519e	0.23	1.31	3.39	0.00	0.03	-0.75	G1 v Control
hsa-miR-4262	0.25	1.49	3.lll23	0.00	0.04	-1.21	G1 v Control
hsa-miR-1289-2	0.29	1.54	3.35	0.00	0.03	-0.86	G1 v Control
hsa-miR-1279	0.29	1.38	3.49	0.00	0.03	-0.45	G1 v Control
hsa-miR-941-2	-0.33	3.08	-3.70	0.00	0.02	0.25	G1 v Control
hsa-miR-532	0.38	3.02	3.63	0.00	0.02	0.01	G1 v Control
hsa-miR-1468	0.46	1.48	3.37	0.00	0.03	-0.79	G1 v Control
hsa-miR-526a-2	0.49	1.66	3.20	0.00	0.05	-1.30	G1 v Control
hsa-miR-548n	0.55	1.73	3.23	0.00	0.04	-1.22	G1 v Control
hsa-miR-335	0.59	2.00	3.45	0.00	0.03	-0.54	G1 v Control
hsa-miR-526a-2	0.61	1.57	3.45	0.00	0.03	-0.55	G1 v Control
hsa-miR-548i-4	0.69	1.84	3.23	0.00	0.04	-1.21	G1 v Control
hsa-miR-302c	0.70	1.52	4.16	0.00	0.01	1.84	G1 v Control
hsa-miR-302c	0.71	1.60	4.24	0.00	0.00	2.13	G1 v Control
hsa-miR-1264	0.72	1.63	3.81	0.00	0.01	0.61	G1 v Control
hsa-miR-548q	0.86	2.19	3.59	0.00	0.02	-0.10	G1 v Control
hsa-miR-519e	0.87	2.16	4.43	0.00	0.00	2.86	G1 v Control
hsa-miR-548f-1	0.87	1.63	4.40	0.00	0.00	2.75	G1 v Control
hsa-miR-520g	0.90	2.60	3.25	0.00	0.04	-1.15	G1 v Control
hsa-miR-517a	0.93	1.95	3.81	0.00	0.01	0.60	G1 v Control
hsa-miR-520h	1.05	2.42	3.75	0.00	0.02	0.42	G1 v Control
hsa-miR-548n	1.10	2.01	4.29	0.00	0.00	2.31	G1 v Control
hsa-miR-522	1.14	1.97	4.20	0.00	0.00	1.99	G1 v Control
hsa-miR-1979	1.14	2.51	3.76	0.00	0.02	0.45	G1 v Control
hsa-miR-520f	1.16	1.83	4.55	0.00	0.00	3.31	G1 v Control
hsa-miR-4275	1.26	2.20	4.03	0.00	0.01	1.38	G1 v Control
hsa-miR-3133	1.38	1.90	4.40	0.00	0.00	2.73	G1 v Control
hsa-miR-520f	1.46	2.13	4.45	0.00	0.00	2.95	G1 v Control
hsa-miR-548i-4	1.47	2.28	3.98	0.00	0.01	1.21	G1 v Control
hsa-miR-3128	1.63	2.56	4.59	0.00	0.00	3.47	G1 v Control
hsa-miR-128-1	0.23	1.26	4.26	0.00	0.02	2.17	G2 v Control
hsa-miR-518f	0.23	1.13	3.83	0.00	0.04	0.70	G2 v Control
hsa-miR-1297	-0.26	1.34	-4.22	0.00	0.02	2.02	G2 v Control
hsa-miR-4320	0.26	1.32	3.80	0.00	0.04	0.60	G2 v Control
hsa-miR-608	0.27	1.39	3.95	0.00	0.04	1.08	G2 v Control

Table 9: Differentially Expressed Stem Loop Pre Micro RNA Species. Differentially Expressed Stem Loop Pre-Micro RNA Species. The ID column is the micro RNA name as defined in the miRBase database[4] accessed via `http://www.mirbase.org/`.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0044260	cellular macromolecule metabolic process	0.00	1.29	1762.86	1920	5427
GO:0031323	regulation of cellular metabolic process	0.00	1.28	1249.63	1388	3847
GO:0006996	organelle organization	0.00	1.33	693.84	805	2136
GO:0006464	cellular protein modification process	0.00	1.32	734.44	847	2261
GO:0036211	protein modification process	0.00	1.32	734.44	847	2261
GO:0080090	regulation of primary metabolic process	0.00	1.26	1217.47	1348	3748
GO:0051252	regulation of RNA metabolic process	0.00	1.30	793.24	907	2442
GO:0019219	regulation of nucleobase-containing compound metabolic process	0.00	1.28	951.43	1072	2929
GO:0051253	negative regulation of RNA metabolic process	0.00	1.52	256.94	328	791
GO:0043412	macromolecule modification	0.00	1.31	757.83	869	2333
GO:2000112	regulation of cellular macromolecule biosynthetic process	0.00	1.29	835.79	950	2573
GO:0010556	regulation of macromolecule biosynthetic process	0.00	1.28	861.45	974	2652
GO:1902679	negative regulation of RNA biosynthetic process	0.00	1.50	248.82	316	766
GO:0010558	negative regulation of macromolecule biosynthetic process	0.00	1.46	288.45	360	888
GO:0006139	nucleobase-containing compound metabolic process	0.00	1.24	1389.95	1517	4279
GO:0006351	transcription, DNA-templated	0.00	1.28	836.77	947	2576
GO:0010629	negative regulation of gene expression	0.00	1.47	273.51	343	842
GO:0007399	nervous system development	0.00	1.34	523.30	615	1611
GO:0045892	negative regulation of transcription, DNA-templated	0.00	1.50	243.62	309	750
GO:0051171	regulation of nitrogen compound metabolic process	0.00	1.26	974.82	1089	3001
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	0.00	1.46	276.43	345	851
GO:0006355	regulation of transcription, DNA-templated	0.00	1.28	759.13	863	2337
GO:1902589	single-organism organelle organization	0.00	1.35	461.59	547	1421
GO:2001141	regulation of RNA biosynthetic process	0.00	1.28	769.20	873	2368
GO:0050794	regulation of cellular process	0.00	1.23	2120.17	2245	6527

Table 10: Biological Process Gene Ontology Enrichment for Stem Loop Pre-Micro RNA Targets. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0016070	RNA metabolic process	0.00	1.25	1019.97	1132	3140
GO:0060255	regulation of macromolecule metabolic process	0.00	1.24	1151.20	1267	3544
GO:0031326	regulation of cellular biosynthetic process	0.00	1.26	898.48	1005	2766
GO:0009889	regulation of biosynthetic process	0.00	1.25	907.58	1013	2794
GO:0043170	macromolecule metabolic process	0.00	1.21	1931.77	2052	5947
GO:0032774	RNA biosynthetic process	0.00	1.25	855.93	958	2635
GO:0010468	regulation of gene expression	0.00	1.24	936.49	1041	2883
GO:0031327	negative regulation of cellular biosynthetic process	0.00	1.40	301.12	366	927
GO:0007010	cytoskeleton organization	0.00	1.45	237.78	296	732
GO:0009890	negative regulation of biosynthetic process	0.00	1.39	304.37	369	937
GO:0030182	neuron differentiation	0.00	1.38	313.79	379	966
GO:0044767	single-organism developmental process	0.00	1.21	1283.73	1393	3952
GO:0045934	negative regulation of nucleobase-containing compound metabolic process	0.00	1.40	282.28	344	869
GO:0034654	nucleobase-containing compound biosynthetic process	0.00	1.23	931.29	1030	2867
GO:0032502	developmental process	0.00	1.21	1296.40	1404	3991
GO:0048699	generation of neurons	0.00	1.36	339.12	405	1044
GO:0051172	negative regulation of nitrogen compound metabolic process	0.00	1.39	287.15	348	884
GO:0046483	heterocycle metabolic process	0.00	1.20	1425.36	1534	4388
GO:0006725	cellular aromatic compound metabolic process	0.00	1.20	1429.58	1538	4401
GO:0048666	neuron development	0.00	1.45	213.41	266	657
GO:0090304	nucleic acid metabolic process	0.00	1.21	1149.58	1252	3539
GO:0010605	negative regulation of macromolecule metabolic process	0.00	1.31	418.71	489	1289
GO:0019222	regulation of metabolic process	0.00	1.20	1381.83	1488	4254
GO:0044267	cellular protein metabolic process	0.00	1.22	925.44	1020	2849
GO:0034645	cellular macromolecule biosynthetic process	0.00	1.21	1088.83	1188	3352

Table 10: Biological Process Gene Ontology Enrichment for Stem Loop Pre-Micro RNA Targets. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0032989	cellular component morphogenesis	0.00	1.35	312.49	373	962
GO:0048869	cellular developmental process	0.00	1.22	853.33	944	2627
GO:0009059	macromolecule biosynthetic process	0.00	1.20	1120.67	1219	3450
GO:0000904	cell morphogenesis involved in differentiation	0.00	1.42	217.31	268	669
GO:0031175	neuron projection development	0.00	1.41	221.86	273	683
GO:0048522	positive regulation of cellular process	0.00	1.21	951.43	1044	2929
GO:0000902	cell morphogenesis	0.00	1.35	296.90	355	914
GO:0048468	cell development	0.00	1.28	465.81	536	1434
GO:0019438	aromatic compound biosynthetic process	0.00	1.21	948.51	1040	2920
GO:0006366	transcription from RNA polymerase II promoter	0.00	1.29	431.38	499	1328
GO:0018130	heterocycle biosynthetic process	0.00	1.21	946.88	1038	2915
GO:0031325	positive regulation of cellular metabolic process	0.00	1.25	579.82	656	1785
GO:0051056	regulation of small GTPase mediated signal transduction	0.00	1.60	110.77	147	341
GO:0048667	cell morphogenesis involved in neuron differentiation	0.00	1.51	145.85	187	449
GO:0010467	gene expression	0.00	1.19	1331.81	1431	4100
GO:0022008	neurogenesis	0.00	1.31	358.29	420	1103
GO:0034641	cellular nitrogen compound metabolic process	0.00	1.18	1478.31	1579	4551
GO:0050789	regulation of biological process	0.00	1.18	2225.74	2326	6852
GO:0000122	negative regulation of transcription from RNA polymerase II promoter	0.00	1.49	154.94	197	477
GO:0010604	positive regulation of macromolecule metabolic process	0.00	1.25	560.01	634	1724
GO:0044271	cellular nitrogen compound biosynthetic process	0.00	1.20	964.10	1054	2968
GO:0009893	positive regulation of metabolic process	0.00	1.24	610.68	687	1880
GO:0046626	regulation of insulin receptor signaling pathway	0.00	4.87	9.74	21	30
GO:0009653	anatomical structure morphogenesis	0.00	1.24	613.28	689	1888
GO:0007275	multicellular organismal development	0.00	1.19	1141.46	1234	3514
GO:0030154	cell differentiation	0.00	1.21	803.31	886	2473

Table 10: Biological Process Gene Ontology Enrichment for Stem Loop Pre-Micro RNA Targets. The ‘GO ID’ is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0048812	neuron projection morphogenesis	0.00	1.44	176.71	220	544
GO:0009966	regulation of signal transduction	0.00	1.24	589.89	663	1816
GO:0023051	regulation of signaling	0.00	1.23	659.08	735	2029
GO:0071840	cellular component organization or biogenesis	0.00	1.18	1297.38	1391	3994
GO:0006468	protein phosphorylation	0.00	1.31	333.28	390	1026
GO:0030036	actin cytoskeleton organization	0.00	1.51	129.28	166	398
GO:1901362	organic cyclic compound biosynthetic process	0.00	1.19	972.87	1059	2995
GO:0031324	negative regulation of cellular metabolic process	0.00	1.27	419.68	482	1292
GO:0048518	positive regulation of biological process	0.00	1.19	1059.92	1148	3263
GO:0016043	cellular component organization	0.00	1.18	1277.56	1370	3933
GO:0065007	biological regulation	0.00	1.17	2349.18	2442	7232
GO:0006357	regulation of transcription from RNA polymerase II promoter	0.00	1.28	381.68	441	1175
GO:0010646	regulation of cell communication	0.00	1.22	661.68	736	2037
GO:0030029	actin filament-based process	0.00	1.47	142.28	180	438
GO:0048856	anatomical structure development	0.00	1.18	1109.95	1196	3417
GO:1901360	organic cyclic compound metabolic process	0.00	1.16	1473.43	1565	4536
GO:0010628	positive regulation of gene expression	0.00	1.29	326.46	380	1005
GO:1900076	regulation of cellular response to insulin stimulus	0.00	3.69	11.69	23	36
GO:0051128	regulation of cellular component organization	0.00	1.26	418.71	478	1289
GO:0009892	negative regulation of metabolic process	0.00	1.25	454.76	516	1400
GO:1900077	negative regulation of cellular response to insulin stimulus	0.00	5.07	7.80	17	24
GO:0006325	chromatin organization	0.00	1.42	158.19	196	487
GO:0045893	positive regulation of transcription, DNA-templated	0.00	1.30	299.82	350	923
GO:0001932	regulation of protein phosphorylation	0.00	1.33	246.22	292	758
GO:0051493	regulation of cytoskeleton organization	0.00	1.58	87.70	116	270
GO:0061564	axon development	0.00	1.44	143.58	179	442

Table 10: Biological Process Gene Ontology Enrichment for Stem Loop Pre-Micro RNA Targets. The ‘GO ID’ is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0030030	cell projection organization	0.00	1.29	298.84	348	920
GO:0016568	chromatin modification	0.00	1.43	139.68	174	430
GO:0051174	regulation of phosphorus metabolic process	0.00	1.24	426.50	483	1313
GO:0032990	cell part morphogenesis	0.00	1.34	216.34	258	666
GO:0070848	response to growth factor	0.00	1.37	186.13	225	573
GO:0007409	axonogenesis	0.00	1.40	155.27	191	478
GO:0032268	regulation of cellular protein metabolic process	0.00	1.25	396.62	451	1221
GO:0046627	negative regulation of insulin receptor signaling pathway	0.00	4.77	7.47	16	23
GO:0006513	protein monoubiquitination	0.00	3.13	12.99	24	40
GO:1901699	cellular response to nitrogen compound	0.00	1.44	129.61	162	399
GO:0022603	regulation of anatomical structure morphogenesis	0.00	1.35	194.25	233	598
GO:0051254	positive regulation of RNA metabolic process	0.00	1.27	316.71	365	975
GO:0070727	cellular macromolecule localization	0.00	1.29	273.83	319	843
GO:0010769	regulation of cell morphogenesis involved in differentiation	0.00	1.64	66.59	90	205
GO:0045935	positive regulation of nucleobase-containing compound metabolic process	0.00	1.25	352.77	403	1086
GO:0042325	regulation of phosphorylation	0.00	1.28	295.60	342	910
GO:0033043	regulation of organelle organization	0.00	1.35	185.48	223	571
GO:0071363	cellular response to growth factor stimulus	0.00	1.36	180.93	218	557
GO:0019220	regulation of phosphate metabolic process	0.00	1.23	422.28	476	1300
GO:0010557	positive regulation of macromolecule biosynthetic process	0.00	1.25	350.49	400	1079
GO:1902531	regulation of intracellular signal transduction	0.00	1.25	358.29	408	1103
GO:1901069	guanosine-containing compound catabolic process	0.00	1.42	127.98	159	394
GO:1902680	positive regulation of RNA biosynthetic process	0.00	1.26	310.54	357	956
GO:0034613	cellular protein localization	0.00	1.28	272.21	316	838
GO:0051173	positive regulation of nitrogen compound metabolic process	0.00	1.24	359.59	409	1107

Table 10: Biological Process Gene Ontology Enrichment for Stem Loop Pre-Micro RNA Targets. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0031399	regulation of protein modification process	0.00	1.26	315.41	362	971
GO:0006184	GTP catabolic process	0.00	1.42	127.33	158	392
GO:0009891	positive regulation of biosynthetic process	0.00	1.24	378.75	429	1166
GO:0048523	negative regulation of cellular process	0.00	1.17	859.50	929	2646
GO:0007264	small GTPase mediated signal transduction	0.00	1.39	146.50	179	451
GO:0048858	cell projection morphogenesis	0.00	1.31	213.41	252	657
GO:0022604	regulation of cell morphogenesis	0.00	1.47	101.67	129	313
GO:0031328	positive regulation of cellular biosynthetic process	0.00	1.24	371.93	421	1145
GO:0032956	regulation of actin cytoskeleton organization	0.00	1.64	60.74	82	187
GO:0006807	nitrogen compound metabolic process	0.00	1.14	1569.58	1649	4832
GO:0033138	positive regulation of peptidyl-serine phosphorylation	0.00	2.54	16.57	28	51
GO:0071310	cellular response to organic substance	0.00	1.21	441.12	493	1358
GO:0071417	cellular response to organonitrogen compound	0.00	1.42	120.19	149	370
GO:0006470	protein dephosphorylation	0.00	1.77	43.20	61	133
GO:0046039	GTP metabolic process	0.00	1.40	130.26	160	401
GO:0045444	fat cell differentiation	0.00	1.81	39.95	57	123
GO:0032535	regulation of cellular component size	0.00	1.61	62.04	83	191
GO:0051276	chromosome organization	0.00	1.31	208.22	245	641
GO:1901068	guanosine-containing compound metabolic process	0.00	1.39	132.21	162	407
GO:0048813	dendrite morphogenesis	0.00	2.09	25.34	39	78
GO:0008064	regulation of actin polymerization or depolymerization	0.00	1.90	33.46	49	103
GO:0016310	phosphorylation	0.00	1.21	420.98	471	1296

Table 10: Biological Process Gene Ontology Enrichment for Stem Loop Pre-Micro RNA Targets. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0005515	protein binding	0.00	1.28	2135.59	2284	6624
GO:0005488	binding	0.00	1.30	3098.28	3187	9610
GO:0019899	enzyme binding	0.00	1.38	336.91	407	1045
GO:0004842	ubiquitin-protein ligase activity	0.00	1.77	74.47	105	231
GO:0097159	organic cyclic compound binding	0.00	1.18	1423.08	1522	4414
GO:0016881	acid-amino acid ligase activity	0.00	1.68	83.18	114	258
GO:1901363	heterocyclic compound binding	0.00	1.18	1404.38	1502	4356
GO:0019787	small conjugating protein ligase activity	0.00	1.71	78.99	109	245
GO:0003676	nucleic acid binding	0.00	1.19	924.97	1007	2869
GO:0003677	DNA binding	0.00	1.23	561.30	629	1741
GO:0019900	kinase binding	0.00	1.49	125.41	160	389
GO:0004721	phosphoprotein phosphatase activity	0.00	1.88	48.04	70	149
GO:0043024	ribosomal small subunit binding	0.00	Inf	2.58	8	8
GO:0043167	ion binding	0.00	1.16	1477.57	1567	4583
GO:0003682	chromatin binding	0.00	1.48	106.07	135	329
GO:0008092	cytoskeletal protein binding	0.00	1.34	195.70	234	607
GO:0016879	ligase activity, forming carbon-nitrogen bonds	0.00	1.51	91.56	118	284
GO:0004674	protein serine/threonine kinase activity	0.00	1.43	121.87	152	378
GO:0005545	1-phosphatidylinositol binding	0.00	4.92	6.45	14	20
GO:0005088	Ras guanyl-nucleotide exchange factor activity	0.00	1.93	34.50	51	107
GO:0016791	phosphatase activity	0.00	1.58	70.28	93	218
GO:0019901	protein kinase binding	0.00	1.43	113.16	141	351
GO:0044822	poly(A) RNA binding	0.00	1.26	293.06	336	909

Table 11: Molecular Function Gene Ontology Enrichment for Stem Loop Pre-Micro RNA Targets. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0005622	intracellular	0.00	1.32	3173.77	3297	9888
GO:0044424	intracellular part	0.00	1.31	3146.81	3270	9804
GO:0005634	nucleus	0.00	1.25	1538.10	1681	4792
GO:0044464	cell part	0.00	1.39	3568.57	3658	11118
GO:0005623	cell	0.00	1.39	3568.89	3658	11119
GO:0043226	organelle	0.00	1.24	2872.38	2990	8949
GO:0044428	nuclear part	0.00	1.27	667.94	760	2081
GO:0043229	intracellular organelle	0.00	1.21	2730.19	2844	8506
GO:0031981	nuclear lumen	0.00	1.28	565.23	650	1761
GO:0043227	membrane-bounded organelle	0.00	1.19	2676.26	2780	8338
GO:0005856	cytoskeleton	0.00	1.27	476.64	547	1485
GO:0043228	non-membrane-bounded organelle	0.00	1.21	837.42	925	2609
GO:0043232	intracellular non-membrane-bounded organelle	0.00	1.21	837.42	925	2609
GO:0043231	intracellular membrane-bounded organelle	0.00	1.17	2474.05	2577	7708
GO:0030054	cell junction	0.00	1.37	223.72	271	697
GO:0005654	nucleoplasm	0.00	1.26	400.25	458	1247
GO:0035770	ribonucleoprotein granule	0.00	2.23	26.32	42	82
GO:0031252	cell leading edge	0.00	1.54	85.06	111	265
GO:0044430	cytoskeletal part	0.00	1.25	335.74	384	1046
GO:0005911	cell-cell junction	0.00	1.52	83.45	108	260
GO:1902494	catalytic complex	0.00	1.30	213.13	251	664
GO:0010494	cytoplasmic stress granule	0.00	3.47	9.31	18	29

Table 12: Cellular Component Gene Ontology Enrichment for Stem Loop Pre-Micro RNA Targets. The ‘GO ID’ is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

4.4 Small Nucleolar Micro RNAs

Differentially expressed small nucleolar RNA species are shown in Table 13. The majority of these differences are associated with the comparison between the hgigh exposure group (Group 1) and the control group (Group 4).

Many of the probe sets are associated with U3 or U13. Unusually for small nucleolar micro RNAs, neither of these micro RNAs are involved in methylation or pseudouridylation; instead both are involved in ribosomal RNA cleavage[11]. There is no obvious clinical relevance of this finding, given current knowledge.

ID	logFC	AveExpr	t	P.Value	adj.P.Val	B	Contrast
U3	0.16	1.18	3.58	0.00	0.03	-0.13	G1 v Control
snoU13	-0.18	1.29	-3.35	0.00	0.05	-0.84	G1 v Control
SNORD78	-0.19	1.21	-3.49	0.00	0.04	-0.41	G1 v Control
snoU13	-0.21	2.25	-3.45	0.00	0.04	-0.52	G1 v Control
SNORD11	0.23	1.23	4.48	0.00	0.01	3.04	G1 v Control
snoU13	0.24	1.68	3.54	0.00	0.04	-0.24	G1 v Control
snoZ40	0.26	1.31	3.86	0.00	0.02	0.80	G1 v Control
SNORA73	0.26	1.32	3.80	0.00	0.02	0.58	G1 v Control
SNORA1	0.27	1.28	3.61	0.00	0.03	-0.03	G1 v Control
SNORD112	0.29	1.31	3.31	0.00	0.05	-0.96	G1 v Control
U3	0.33	1.36	4.21	0.00	0.01	2.00	G1 v Control
SNORA43	0.33	1.38	3.65	0.00	0.03	0.11	G1 v Control
SNORD112	0.36	1.55	3.32	0.00	0.05	-0.91	G1 v Control
snoU13	0.37	1.67	3.45	0.00	0.04	-0.52	G1 v Control
snoU13	0.37	2.00	3.31	0.00	0.05	-0.94	G1 v Control
SNORD112	0.38	1.32	4.00	0.00	0.02	1.27	G1 v Control
SNORD28	0.38	1.51	4.02	0.00	0.02	1.34	G1 v Control
SNORA81	0.40	1.45	3.33	0.00	0.05	-0.89	G1 v Control
snoU13	0.45	1.48	3.72	0.00	0.03	0.33	G1 v Control
SNORA31	0.45	1.67	3.94	0.00	0.02	1.07	G1 v Control
SNORA41	0.46	1.79	3.65	0.00	0.03	0.10	G1 v Control
SNORD113	0.46	1.39	3.84	0.00	0.02	0.73	G1 v Control
SNORD112	0.52	1.45	4.75	0.00	0.01	4.09	G1 v Control
SNORA1	0.52	1.49	3.51	0.00	0.04	-0.36	G1 v Control
SNORA31	0.61	1.72	3.50	0.00	0.04	-0.36	G1 v Control
U3	0.64	2.22	3.75	0.00	0.03	0.42	G1 v Control
U3	0.72	1.87	3.45	0.00	0.04	-0.53	G1 v Control
U3	0.72	1.73	3.85	0.00	0.02	0.77	G1 v Control
U3	0.77	3.70	3.36	0.00	0.05	-0.79	G1 v Control
snoU13	0.95	6.36	3.38	0.00	0.05	-0.73	G1 v Control
U3	0.98	3.25	3.43	0.00	0.04	-0.60	G1 v Control
SNORA1	1.07	2.03	4.48	0.00	0.01	3.03	G1 v Control
snosnR66	1.08	2.63	3.42	0.00	0.04	-0.63	G1 v Control
SNORA43	1.28	2.20	3.70	0.00	0.03	0.25	G1 v Control
U3	1.28	3.25	3.54	0.00	0.04	-0.25	G1 v Control
U3	1.29	3.03	3.85	0.00	0.02	0.77	G1 v Control
SNORA43	1.33	2.28	3.92	0.00	0.02	1.00	G1 v Control
SNORD50	1.62	2.09	4.36	0.00	0.01	2.57	G1 v Control
U3	0.32	1.54	4.07	0.00	0.04	1.44	G2 v Control
snoU13	0.42	1.98	4.09	0.00	0.04	1.51	G2 v Control
snoU13	0.44	1.45	4.32	0.00	0.04	2.30	G2 v Control
SNORD11	0.31	1.23	5.25	0.00	0.00	5.81	Group 3 v Control
ACA64	0.37	2.54	4.08	0.00	0.04	1.46	Group 3 v Control
U8	0.43	1.76	4.09	0.00	0.04	1.48	Group 3 v Control
SNORA70	0.44	2.45	3.96	0.00	0.04	1.03	Group 3 v Control

Table 13: Differentially Expressed Small Nucleolar Micro RNA Species. The ID column is the micro RNA name as defined in the miRBase database[4] accessed via `http://www.mirbase.org/`.

5 Relationship With Health Scores

The empirical Bayes analysis was repeated using a model involving regression of each micro RNA species on the health score. Three mature micro RNA species were significant following Benjamini Hochberg adjustment to a 0.05 false discovery rate. No stem-loop or small nucleolar micro RNA species were statistically significant. The statistically significant mature micro RNA species are shown in Table 14. Note that the log fold change is for a unit change in total health questionnaire score; the inter-quartile range for the total score is approximately 20.

ID	logFC	AveExpr	t	P.Value	adj.P.Val	B
hsa-miR-103	-0.01	11.95	-4.33	0.00	0.02	0.46
hsa-miR-107	-0.01	11.59	-4.18	0.00	0.02	-0.11
hsa-miR-200b	0.01	1.38	4.21	0.00	0.02	-0.01

Table 14: Mature Micro RNA Species With Statistically Significant Relationship With Health Score Following Benjamini Hochberg Adjustment. The ID column is the micro RNA name as defined in the miRBase database[4] accessed via `http://www.mirbase.org/`.

The three statistically significant miRNA species from Table 14 are associated with 4003 mRNA targets. Gene Ontology enrichment analyses for the target mRNAs are shown in Tables 15, 16, and 17 for Biological Process, Molecular Function and Cellular Component domains respectively.

We note the for the Biological process domain there are a number of terms associated with RNA metabolic processes and transcription control, and also terms associated with central nervous system development and function. A primary ontology association with CNS development is not necessarily inconsistent with a role in CNS function and cognition in adulthood.

The Molecular Function domain shows enrichment in several ontology terms associated with chromatin and transcription factor binding. There are no obvious groupings in the Cellular Component domain.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0048856	anatomical structure development	0.00	1.32	954.94	1092	3452
GO:0032502	developmental process	0.00	1.30	1115.38	1256	4032
GO:0044767	single-organism developmental process	0.00	1.30	1104.32	1243	3992
GO:0009653	anatomical structure morphogenesis	0.00	1.38	526.99	635	1905
GO:0022008	neurogenesis	0.00	1.49	307.34	392	1111
GO:0007399	nervous system development	0.00	1.40	448.97	548	1623
GO:0048468	cell development	0.00	1.42	400.01	494	1446
GO:0031323	regulation of cellular metabolic process	0.00	1.29	1075.82	1208	3889
GO:0030182	neuron differentiation	0.00	1.51	269.16	348	973
GO:0060255	regulation of macromolecule metabolic process	0.00	1.29	990.34	1119	3580
GO:0048699	generation of neurons	0.00	1.49	291.02	372	1052
GO:0048812	neuron projection morphogenesis	0.00	1.69	151.32	211	547
GO:0007275	multicellular organismal development	0.00	1.28	982.05	1107	3550
GO:0006464	cellular protein modification process	0.00	1.33	631.28	739	2282
GO:0036211	protein modification process	0.00	1.33	631.28	739	2282
GO:0010628	positive regulation of gene expression	0.00	1.48	279.68	357	1011
GO:0050794	regulation of cellular process	0.00	1.26	1823.29	1955	6591
GO:0003007	heart morphogenesis	0.00	2.41	45.92	79	166
GO:0032990	cell part morphogenesis	0.00	1.59	185.07	248	669
GO:0045893	positive regulation of transcription, DNA-templated	0.00	1.49	256.72	329	928
GO:0048858	cell projection morphogenesis	0.00	1.58	182.58	244	660
GO:0000904	cell morphogenesis involved in differentiation	0.00	1.57	187.00	249	676
GO:0019222	regulation of metabolic process	0.00	1.25	1189.24	1313	4299
GO:0043412	macromolecule modification	0.00	1.31	652.02	756	2357
GO:0048731	system development	0.00	1.28	818.83	931	2960

Table 15: Biological Process Gene Ontology Enrichment for Mature Micro RNA Targets Relationship With Health Score. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0080090	regulation of primary metabolic process	0.00	1.26	1047.61	1167	3787
GO:0010604	positive regulation of macromolecule metabolic process	0.00	1.34	479.96	571	1735
GO:0007507	heart development	0.00	1.80	99.59	145	360
GO:0051252	regulation of RNA metabolic process	0.00	1.30	681.35	785	2463
GO:0031175	neuron projection development	0.00	1.54	190.32	251	688
GO:0050789	regulation of biological process	0.00	1.25	1914.57	2036	6921
GO:0048869	cellular developmental process	0.00	1.28	734.74	840	2656
GO:0032989	cellular component morphogenesis	0.00	1.45	268.06	338	969
GO:0009893	positive regulation of metabolic process	0.00	1.32	523.94	616	1894
GO:0051254	positive regulation of RNA metabolic process	0.00	1.44	271.10	341	980
GO:0048666	neuron development	0.00	1.55	181.75	240	657
GO:0048667	cell morphogenesis involved in neuron differentiation	0.00	1.68	124.21	173	449
GO:0030154	cell differentiation	0.00	1.28	692.13	793	2502
GO:0000902	cell morphogenesis	0.00	1.45	254.78	322	921
GO:0019219	regulation of nucleobase-containing compound metabolic process	0.00	1.27	818.56	925	2959
GO:0010468	regulation of gene expression	0.00	1.27	804.17	910	2907
GO:0065007	biological regulation	0.00	1.25	2023.01	2139	7313
GO:2000112	regulation of cellular macromolecule biosynthetic process	0.00	1.28	718.14	820	2596
GO:0006355	regulation of transcription, DNA-templated	0.00	1.29	652.02	750	2357
GO:0031325	positive regulation of cellular metabolic process	0.00	1.32	497.39	585	1798
GO:0007409	axonogenesis	0.00	1.63	132.78	182	480
GO:2001141	regulation of RNA biosynthetic process	0.00	1.28	660.60	758	2388
GO:0044267	cellular protein metabolic process	0.00	1.26	795.32	899	2875
GO:0061564	axon development	0.00	1.65	122.27	169	442
GO:1902680	positive regulation of RNA biosynthetic process	0.00	1.42	265.84	332	961
GO:0044260	cellular macromolecule metabolic process	0.00	1.22	1518.71	1635	5490

Table 15: Biological Process Gene Ontology Enrichment for Mature Micro RNA Targets Relationship With Health Score. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0009887	organ morphogenesis	0.00	1.49	197.52	255	714
GO:0010605	negative regulation of macromolecule metabolic process	0.00	1.36	359.35	434	1299
GO:0006357	regulation of transcription from RNA polymerase II promoter	0.00	1.37	326.98	398	1182
GO:0003205	cardiac chamber development	0.00	2.64	28.77	52	104
GO:0031326	regulation of cellular biosynthetic process	0.00	1.26	772.36	871	2792
GO:0051171	regulation of nitrogen compound metabolic process	0.00	1.25	838.75	940	3032
GO:0007165	signal transduction	0.00	1.23	1024.09	1131	3702
GO:0009790	embryo development	0.00	1.44	228.78	289	827
GO:0010556	regulation of macromolecule biosynthetic process	0.00	1.26	740.27	837	2676
GO:0072358	cardiovascular system development	0.00	1.51	177.60	231	642
GO:0072359	circulatory system development	0.00	1.51	177.60	231	642
GO:0048566	embryonic digestive tract development	0.00	5.79	8.85	22	32
GO:0003279	cardiac septum development	0.00	3.62	15.77	33	57
GO:0009889	regulation of biosynthetic process	0.00	1.25	780.10	877	2820
GO:0023052	signaling	0.00	1.22	1136.68	1244	4109
GO:0044700	single organism signaling	0.00	1.22	1136.68	1244	4109
GO:0048598	embryonic morphogenesis	0.00	1.60	124.76	169	451
GO:0045935	positive regulation of nucleobase-containing compound metabolic process	0.00	1.37	302.36	368	1093
GO:0048646	anatomical structure formation involved in morphogenesis	0.00	1.44	211.07	267	763
GO:0030030	cell projection organization	0.00	1.40	256.99	318	929
GO:0051173	positive regulation of nitrogen compound metabolic process	0.00	1.36	308.17	374	1114
GO:0009966	regulation of signal transduction	0.00	1.29	506.51	587	1831
GO:0032268	regulation of cellular protein metabolic process	0.00	1.34	341.36	409	1234
GO:0007154	cell communication	0.00	1.21	1152.17	1255	4165
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	0.00	1.41	236.80	294	856

Table 15: Biological Process Gene Ontology Enrichment for Mature Micro RNA Targets Relationship With Health Score. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0048518	positive regulation of biological process	0.00	1.22	910.67	1007	3292
GO:0001944	vasculature development	0.00	1.55	135.00	179	488
GO:0043009	chordate embryonic development	0.00	1.53	139.42	184	504
GO:0031328	positive regulation of cellular biosynthetic process	0.00	1.34	318.96	383	1153
GO:0006351	transcription, DNA-templated	0.00	1.24	719.52	807	2601
GO:0010557	positive regulation of macromolecule biosynthetic process	0.00	1.35	300.42	362	1086
GO:0001568	blood vessel development	0.00	1.61	104.84	143	379
GO:0031324	negative regulation of cellular metabolic process	0.00	1.32	360.73	427	1304
GO:0007417	central nervous system development	0.00	1.44	184.79	234	668
GO:0051603	proteolysis involved in cellular protein catabolic process	0.00	1.60	107.61	146	389
GO:0048562	embryonic organ morphogenesis	0.00	1.81	64.73	95	234
GO:0043632	modification-dependent macromolecule catabolic process	0.00	1.61	102.63	140	371
GO:0048523	negative regulation of cellular process	0.00	1.23	739.16	826	2672
GO:0009792	embryo development ending in birth or egg hatching	0.00	1.51	140.81	184	509
GO:0048522	positive regulation of cellular process	0.00	1.22	818.00	907	2957
GO:0009891	positive regulation of biosynthetic process	0.00	1.33	324.77	387	1174
GO:0006366	transcription from RNA polymerase II promoter	0.00	1.30	370.41	436	1339
GO:0010629	negative regulation of gene expression	0.00	1.38	234.31	288	847
GO:0048519	negative regulation of biological process	0.00	1.22	802.23	890	2900
GO:0010558	negative regulation of macromolecule biosynthetic process	0.00	1.37	247.31	302	894
GO:0010646	regulation of cell communication	0.00	1.25	567.93	645	2053
GO:0019941	modification-dependent protein catabolic process	0.00	1.59	101.80	138	368
GO:0045892	negative regulation of transcription, DNA-templated	0.00	1.40	208.86	259	755
GO:0023051	regulation of signaling	0.00	1.25	565.71	642	2045
GO:0030900	forebrain development	0.00	1.66	84.10	117	304
GO:0010467	gene expression	0.00	1.19	1146.09	1241	4143

Table 15: Biological Process Gene Ontology Enrichment for Mature Micro RNA Targets Relationship With Health Score. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0030163	protein catabolic process	0.00	1.49	138.59	180	501
GO:0009892	negative regulation of metabolic process	0.00	1.29	390.61	456	1412
GO:0048589	developmental growth	0.00	1.71	73.31	104	265
GO:0006468	protein phosphorylation	0.00	1.33	286.87	344	1037
GO:0051253	negative regulation of RNA metabolic process	0.00	1.38	220.20	271	796
GO:0006511	ubiquitin-dependent protein catabolic process	0.00	1.58	101.52	137	367
GO:0051716	cellular response to stimulus	0.00	1.19	1234.33	1329	4462
GO:0048568	embryonic organ development	0.00	1.60	94.88	129	343
GO:0045934	negative regulation of nucleobase-containing compound metabolic process	0.00	1.36	241.78	294	874
GO:0031327	negative regulation of cellular biosynthetic process	0.00	1.35	258.37	312	934
GO:0035264	multicellular organism growth	0.00	2.24	30.15	50	109
GO:0032269	negative regulation of cellular protein metabolic process	0.00	1.53	115.91	153	419
GO:0007411	axon guidance	0.00	1.60	92.67	126	335
GO:0097485	neuron projection guidance	0.00	1.60	92.67	126	335
GO:1902679	negative regulation of RNA biosynthetic process	0.00	1.38	213.28	262	771
GO:0001933	negative regulation of protein phosphorylation	0.00	1.79	57.54	84	208
GO:0009890	negative regulation of biosynthetic process	0.00	1.34	261.14	314	944
GO:0003206	cardiac chamber morphogenesis	0.00	2.41	24.34	42	88
GO:0032774	RNA biosynthetic process	0.00	1.21	736.40	816	2662
GO:0043170	macromolecule metabolic process	0.00	1.18	1664.50	1758	6017
GO:0044257	cellular protein catabolic process	0.00	1.57	96.82	130	350
GO:0007167	enzyme linked receptor protein signaling pathway	0.00	1.35	232.92	282	842
GO:0003281	ventricular septum development	0.00	3.56	11.07	23	40
GO:0031076	embryonic camera-type eye development	0.00	4.04	9.13	20	33
GO:0060485	mesenchyme development	0.00	2.06	34.03	54	123

Table 15: Biological Process Gene Ontology Enrichment for Mature Micro RNA Targets Relationship With Health Score. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0051172	negative regulation of nitrogen compound metabolic process	0.00	1.33	246.20	296	890
GO:0038095	Fc-epsilon receptor signaling pathway	0.00	1.92	42.05	64	152
GO:0048813	dendrite morphogenesis	0.00	2.44	21.85	38	79
GO:0003231	cardiac ventricle development	0.00	2.44	21.85	38	79
GO:2000026	regulation of multicellular organismal development	0.00	1.30	298.21	352	1078
GO:0016043	cellular component organization	0.00	1.18	1100.44	1187	3978
GO:0035556	intracellular signal transduction	0.00	1.24	494.62	561	1788
GO:0021915	neural tube development	0.00	2.08	32.64	52	118
GO:0050793	regulation of developmental process	0.00	1.26	388.95	448	1406
GO:0007610	behavior	0.00	1.45	132.51	169	479
GO:0002009	morphogenesis of an epithelium	0.00	1.54	95.71	127	346
GO:0040012	regulation of locomotion	0.00	1.45	126.70	162	458
GO:0032879	regulation of localization	0.00	1.26	372.90	430	1348
GO:0044707	single-multicellular organism process	0.00	1.17	1243.46	1329	4495
GO:2000027	regulation of organ morphogenesis	0.00	1.91	39.56	60	143
GO:0048844	artery morphogenesis	0.00	3.15	12.17	24	44
GO:0007420	brain development	0.00	1.43	141.08	178	510
GO:0006469	negative regulation of protein kinase activity	0.00	1.87	42.05	63	152
GO:0010608	posttranscriptional regulation of gene expression	0.00	1.52	97.93	129	354
GO:0034654	nucleobase-containing compound biosynthetic process	0.00	1.19	801.13	877	2896
GO:0045595	regulation of cell differentiation	0.00	1.29	282.72	333	1022
GO:0071840	cellular component organization or biogenesis	0.00	1.17	1117.87	1201	4041
GO:0033673	negative regulation of kinase activity	0.00	1.82	45.37	67	164
GO:0045944	positive regulation of transcription from RNA polymerase II promoter	0.00	1.37	181.19	222	655
GO:0044708	single-organism behavior	0.00	1.59	77.46	105	280
GO:0040007	growth	0.00	1.35	193.09	235	698

Table 15: Biological Process Gene Ontology Enrichment for Mature Micro RNA Targets Relationship With Health Score. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0001947	heart looping	0.00	3.18	11.62	23	42
GO:0055024	regulation of cardiac muscle tissue development	0.00	3.18	11.62	23	42
GO:0050767	regulation of neurogenesis	0.00	1.46	115.91	149	419
GO:0048513	organ development	0.00	1.21	571.80	638	2067
GO:0048729	tissue morphogenesis	0.00	1.46	118.68	152	429
GO:0016310	phosphorylation	0.00	1.26	362.11	417	1309
GO:2000145	regulation of cell motility	0.00	1.46	117.02	150	423
GO:0023057	negative regulation of signaling	0.00	1.34	203.60	246	736
GO:0009968	negative regulation of signal transduction	0.00	1.35	194.47	236	703
GO:0003157	endocardium development	0.00	20.97	2.49	8	9
GO:0044265	cellular macromolecule catabolic process	0.00	1.37	173.73	213	628
GO:0040011	locomotion	0.00	1.27	318.68	370	1152
GO:0000122	negative regulation of transcription from RNA polymerase II promoter	0.00	1.42	133.34	168	482
GO:0018130	heterocycle biosynthetic process	0.00	1.18	815.24	889	2947
GO:0033036	macromolecule localization	0.00	1.22	466.96	527	1688
GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	0.00	1.37	167.64	206	606
GO:0019538	protein metabolic process	0.00	1.17	935.29	1012	3381
GO:0006935	chemotaxis	0.00	1.40	146.89	183	531
GO:0042330	taxis	0.00	1.40	146.89	183	531
GO:0007166	cell surface receptor signaling pathway	0.00	1.20	590.33	656	2134
GO:0010648	negative regulation of cell communication	0.00	1.33	204.15	246	738
GO:0051270	regulation of cellular component movement	0.00	1.42	133.61	168	483
GO:0048736	appendage development	0.00	1.87	38.73	58	140
GO:0060173	limb development	0.00	1.87	38.73	58	140
GO:0044271	cellular nitrogen compound biosynthetic process	0.00	1.18	829.62	903	2999
GO:0038093	Fc receptor signaling pathway	0.00	1.68	56.16	79	203

Table 15: Biological Process Gene Ontology Enrichment for Mature Micro RNA Targets Relationship With Health Score. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0032501	multicellular organismal process	0.00	1.16	1285.51	1367	4647
GO:0003151	outflow tract morphogenesis	0.00	3.02	11.90	23	43
GO:0060602	branch elongation of an epithelium	0.00	7.21	4.15	11	15
GO:0051246	regulation of protein metabolic process	0.00	1.24	383.41	438	1386
GO:0048596	embryonic camera-type eye morphogenesis	0.00	4.20	7.19	16	26
GO:0019438	aromatic compound biosynthetic process	0.00	1.18	816.34	889	2951
GO:0016070	RNA metabolic process	0.00	1.17	877.76	952	3173
GO:0048514	blood vessel morphogenesis	0.00	1.52	88.80	117	321
GO:0035295	tube development	0.00	1.46	108.99	140	394
GO:0030334	regulation of cell migration	0.00	1.45	111.76	143	404
GO:0007423	sensory organ development	0.00	1.46	109.27	140	395
GO:0043254	regulation of protein complex assembly	0.00	1.68	54.77	77	198
GO:0048583	regulation of response to stimulus	0.00	1.19	649.81	716	2349
GO:0032880	regulation of protein localization	0.00	1.45	111.21	142	402
GO:0060411	cardiac septum morphogenesis	0.00	3.04	11.34	22	41
GO:0048638	regulation of developmental growth	0.00	1.94	32.64	50	118
GO:0001701	in utero embryonic development	0.00	1.52	84.10	111	304
GO:0009059	macromolecule biosynthetic process	0.00	1.16	964.62	1039	3487
GO:0060840	artery development	0.00	2.74	13.55	25	49
GO:0060839	endothelial cell fate commitment	0.00	Inf	1.66	6	6
GO:0034645	cellular macromolecule biosynthetic process	0.00	1.16	937.51	1011	3389
GO:0021537	telencephalon development	0.00	1.78	42.60	62	154
GO:0048864	stem cell development	0.00	1.73	46.75	67	169
GO:0048863	stem cell differentiation	0.00	1.57	71.37	96	258
GO:0051248	negative regulation of protein metabolic process	0.00	1.40	131.40	164	475
GO:0006996	organelle organization	0.00	1.19	597.53	660	2160

Table 15: Biological Process Gene Ontology Enrichment for Mature Micro RNA Targets Relationship With Health Score. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0031399	regulation of protein modification process	0.00	1.27	271.93	317	983
GO:0016358	dendrite development	0.00	1.86	35.41	53	128
GO:0070848	response to growth factor	0.00	1.36	158.79	194	574
GO:0003143	embryonic heart tube morphogenesis	0.00	2.74	13.00	24	47
GO:0045639	positive regulation of myeloid cell differentiation	0.00	2.39	17.43	30	63
GO:0048557	embryonic digestive tract morphogenesis	0.00	5.25	4.98	12	18
GO:0060045	positive regulation of cardiac muscle cell proliferation	0.00	5.25	4.98	12	18
GO:0060317	cardiac epithelial to mesenchymal transition	0.00	5.25	4.98	12	18
GO:0009799	specification of symmetry	0.00	2.12	22.96	37	83
GO:0003197	endocardial cushion development	0.00	3.94	6.92	15	25
GO:0031054	pre-miRNA processing	0.00	18.34	2.21	7	8
GO:0016331	morphogenesis of embryonic epithelium	0.00	1.89	31.81	48	115
GO:0043620	regulation of DNA-templated transcription in response to stress	0.00	2.90	11.07	21	40
GO:0048048	embryonic eye morphogenesis	0.00	3.43	8.30	17	30
GO:0071363	cellular response to growth factor stimulus	0.00	1.35	154.36	188	558
GO:0043618	regulation of transcription from RNA polymerase II promoter in response to stress	0.00	3.12	9.68	19	35
GO:0001932	regulation of protein phosphorylation	0.00	1.30	212.18	251	767
GO:0045664	regulation of neuron differentiation	0.00	1.45	95.16	122	344
GO:0060251	regulation of glial cell proliferation	0.00	6.55	3.87	10	14
GO:0034504	protein localization to nucleus	0.00	1.56	64.73	87	234
GO:0014887	cardiac muscle adaptation	0.00	7.86	3.32	9	12
GO:0045948	positive regulation of translational initiation	0.00	7.86	3.32	9	12
GO:0044699	single-organism process	0.00	1.19	2623.58	2681	9484
GO:0042325	regulation of phosphorylation	0.00	1.26	254.50	296	920
GO:0060284	regulation of cell development	0.00	1.35	145.79	178	527

Table 15: Biological Process Gene Ontology Enrichment for Mature Micro RNA Targets Relationship With Health Score. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0009987	cellular process	0.00	1.25	2875.32	2920	10394
GO:0048585	negative regulation of response to stimulus	0.00	1.28	230.43	270	833
GO:0014031	mesenchymal cell development	0.00	2.01	24.90	39	90
GO:0006928	cellular component movement	0.00	1.22	357.13	405	1291
GO:1902589	single-organism organelle organization	0.00	1.21	396.97	447	1435
GO:0007389	pattern specification process	0.00	1.42	105.40	133	381

Table 15: Biological Process Gene Ontology Enrichment for Mature Micro RNA Targets Relationship With Health Score. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0005515	protein binding	0.00	1.30	1838.01	1986	6695
GO:0005488	binding	0.00	1.41	2668.48	2773	9720
GO:0016881	acid-amino acid ligase activity	0.00	1.71	71.93	102	262
GO:0004842	ubiquitin-protein ligase activity	0.00	1.75	64.52	93	235
GO:0019787	small conjugating protein ligase activity	0.00	1.71	68.36	97	249
GO:0000982	RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity	0.00	2.34	25.26	43	92
GO:0016879	ligase activity, forming carbon-nitrogen bonds	0.00	1.61	79.07	108	288
GO:0044822	poly(A) RNA binding	0.00	1.31	252.57	301	920
GO:0003676	nucleic acid binding	0.00	1.19	796.43	872	2901
GO:0003682	chromatin binding	0.00	1.54	91.15	121	332
GO:1901363	heterocyclic compound binding	0.00	1.16	1209.33	1290	4405

Table 16: Molecular Function Gene Ontology Enrichment for Mature Micro RNA Targets Relationship With Health Score. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0008270	zinc ion binding	0.00	1.28	267.95	314	976
GO:0097159	organic cyclic compound binding	0.00	1.15	1225.25	1304	4463
GO:0046872	metal ion binding	0.00	1.17	842.27	913	3068
GO:0043169	cation binding	0.00	1.16	856.55	927	3120
GO:0000987	core promoter proximal region sequence-specific DNA binding	0.00	2.66	13.73	25	50
GO:0003723	RNA binding	0.00	1.24	332.74	381	1212
GO:0000981	sequence-specific DNA binding RNA polymerase II transcription factor activity	0.00	1.58	64.24	87	234
GO:0043167	ion binding	0.00	1.15	1274.12	1350	4641
GO:0046332	SMAD binding	0.00	2.47	15.37	27	56

Table 16: Molecular Function Gene Ontology Enrichment for Mature Micro RNA Targets Relationship With Health Score. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0044451	nucleoplasm part	0.00	1.49	154.66	201	571
GO:0044464	cell part	0.00	1.31	3048.52	3116	11255
GO:0005623	cell	0.00	1.31	3048.79	3116	11256
GO:0005634	nucleus	0.00	1.17	1313.94	1410	4851
GO:0005815	microtubule organizing center	0.00	1.51	114.84	151	424
GO:0005622	intracellular	0.00	1.21	2710.49	2790	10007
GO:0005829	cytosol	0.00	1.21	588.58	658	2173
GO:0044424	intracellular part	0.00	1.19	2687.20	2762	9921

Table 17: Cellular Component Gene Ontology Enrichment for Mature Micro RNA Targets Relationship With Health Score. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

GO ID	GO Term	P value	Odds Ratio	Expected Count	Observed Count	Size
GO:0043229	intracellular organelle	0.00	1.17	2331.02	2415	8606
GO:0005813	centrosome	0.00	1.51	87.49	115	323
GO:0045202	synapse	0.00	1.40	129.47	162	478
GO:0043234	protein complex	0.00	1.17	816.10	887	3013
GO:0005654	nucleoplasm	0.00	1.24	341.82	391	1262
GO:0016604	nuclear body	0.00	1.52	73.94	98	273
GO:0005768	endosome	0.00	1.36	143.28	176	529
GO:0043231	intracellular membrane-bounded organelle	0.00	1.14	2112.43	2190	7799

Table 17: Cellular Component Gene Ontology Enrichment for Mature Micro RNA Targets Relationship With Health Score. The 'GO ID' is the unique identifier in the Gene Ontology Consortium ontology database, and may be searched via the AMI GO interface `http://amigo.geneontology.org/amigo`. The GO Term is the descriptive text provided by the Gene ontology Consortium.

6 Summary and Conclusions

1. There is clear evidence of differential expression of each family of micro RNAs between the high exposure group and the controls.
2. Gene ontology enrichment analysis for the differential mature micro RNAs and stem loop pre-micro RNAs revealed consistent findings. Both types of micro RNA showed enrichment for neurological development processes and for RNA metabolism.
3. Only the mature micro RNAs showed a statistically significant relationship with the health score. Ontology enrichment for the micro RNA relationships with health score showed enrichment for terms associated with neurological development.

These ontology enrichment findings were not present in gene expression analysis for messenger RNAs. Their clinical relevance is uncertain.

References

[1] Quality Assessment of Exon Arrays Revision Date: 2005-09-27. Whitepaper, Affymetrix Inc., 3420 Central Expressway, Santa Clara, CA 95051, April 2007.

[2] D. Betel, M. Wilson, A. Gabow, D. S. Marks, and C. Sander. The microRNA.org resource: targets and expression. *Nucleic Acids Research*, 36(suppl 1):D149–D153, 2008.

[3] B. Everitt, Sabine Landau, and Morven Leese. *Cluster Analysis*. Edward Arnold, 4 edition, May 2001.

[4] S. Griffiths-Jones, H. K. Saini, S. van Dongen, and A. J. Enright. mirbase: tools for microrna genomics. *Nucleic Acids Research*, 36(Database):D154D158, Dec 2007.

[5] J. A. Hartigan. Consistency of Single Linkage for High-Density Clusters. *Journal of the American Statistical Association*, 76(374):388–394, 1981.

[6] P. Huber. *Robust Statistics*. John Wiley and Sons, 1981.

[7] R. Irizarry, B. Hobbs, F. Collin, Y. Beazer-Barclay, K. Antonellis, U. Scherf, and T. Speed. Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data. *Biostatistics*, 4:249 264, 2003.

[8] B. John, A. J. Enright, A. Aravin, T. Tuschl, C. Sander, and D. S. Marks. Human MicroRNA Targets. *PLoS Biology*, 2(11):e363, 2004.

[9] I. Jolliffe. *Principal Components Analysis*. Springer-Verlag, 1986.

[10] A. Kasprzyk. BioMart: driving a paradigm change in biological data management. *Database*, 2011, 2011.

[11] A. M. Kiss, B. E. Jady, E. Bertrand, and T. Kiss. Human Box H/ACA Pseudouridylation Guide RNA Machinery. *Molecular and Cellular Biology*, 24(13):5797–5807, Jul 2004.

[12] W. J. Krzanowski. Principles of Multivariate Analysis: A User's Perspective. Number 3 in Oxford Statistical Science Series, 1988.

[13] H. E. Lockstone. Exon array data analysis using Affymetrix power tools and R statistical software. *Briefings in Bioinformatics*, 2011.

[14] I. Lonnstedt and T. Speed. Replicated Microarray Data. *Statistica Sinica*, pages 31–46, 2002.

[15] J. M. Lucas and M. S. Saccucci. Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements. *Technometrics*, 32(1):1–12, 1990.

[16] P. B. Robinson and T. Y. Ho. Average Run Lengths of Geometric Moving Average Charts by Numerical Methods. *Technometrics*, 20(1):85–93, 1978.

[17] B. Silverman. *Density estimation for statistics and data analysis*. Chapman and Hall, London, 1986.

[18] G. K. Smyth. Limma: linear models for microarray data. In R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, and W. Huber, editors, *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, pages 397–420, New York, 2005. Springer.

[19] M. P. Terns and R. M. Terns. Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin. *Gene expression*, 10(1-2):1–2, 2002.

[20] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, 57:289–300, 1995.

Appendix 5
Report from Emphron on Bioinformatic analysis of mitochondrial DNA array data from leucocytes of exposed workers and controls.

Jet Fuel and Solvents Exposure Project

Mitochondrial DNA

Emphron Informatics Pty Ltd **www.emphron.com**

Report Number: 2012.020
Version Number: 1.0

Client: Jet Fuel and Solvents Project
Date: July 23, 2014
Author: Mervyn Thomas
Email: mervyn.thomas@emphron.com

Signature: .....

1 Introduction

1.1 Background and Objectives

The Jet Fuel and Solvents Exposure study has been set up to investigate physiological and other differences between ex Air Force employees who were exposed to aviation fuel during their service. Subjects were stratified into four exposure groups, where Group 4 is a control group (representing no known exposure) and Group 1 is the highest level of exposure.

Data are available from several genomic and proteomic modalities, including: cytogenetics data, High Throughput mRNA gene expression data, micro RNA expression data, serum analysis by LCMS and full blood counts. This report is concerned with the analysis of mitochondrial DNA variants.

1.2 Available Data

Mitochondrial DNA variant data are available from 245 subjects. 82 of these subjects were in the high exposure group; 43 in the medium exposure group; 43 in the low exposure group and 77 in the control group.

There were 770 variants in total. Most variants occurred in a small proportion of the samples. 434 of the variants occurred in only one sample. Only 12 variants were found in more than 20% but less than 80% of the samples. These 12 variants are the major source of diversity.

2 Exploratory Analysis

Variants which occured in only 1 or 2 samples were dropped. A total of 224 variants remained for analsyis.

A Bray Curtis[1] dissimilarity matrix was calculated between the samples, and the dissimilarity matrix was used to produce a two dimensional metric scaling solution (principal coordinates[3]). Observations were plotted in the space of the principal coordinates, and coloured by exposure group.

Figure 1 shows the scaling solution. There is no relationship between the coordinates nd exposure group. There do appear to be distinct groups in the data. This, however, is an artefact of the definition of the scaling dimensions.

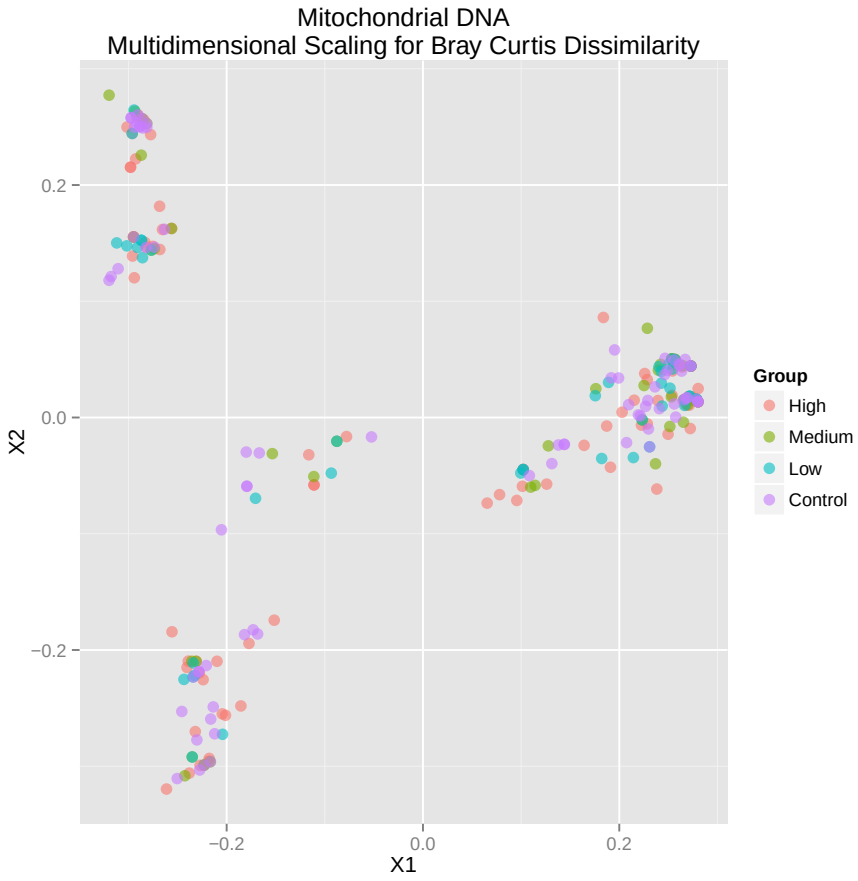


Figure 1: Sample Points in Space of First Two Principal Coordinates of Bray Curtis Dissimilariy

Table 1 shows the correlation between incidence of particular variants and the scaling dimensions. Only variants with a correlation of at least 0.8 on either dimension are displayed. The first scaling dimension is dominated by 8 variants. The second dimension is dominated by the variants at positions 11467,12308 and 12372 .

The discrete groups in the scaling plot are generated by the small number of possible values for this limited set of variants.

Table1 also shows the prevalence for each of the selected variants. Prevalence for these variants is between 19% and 53%.

	Correlation with:		
Variant	Dim. 1	Dim. 2	Prevalence
X73A2G	-92	-6	50
X2706A2G	-91	-11	53
X7028C2T	-91	-11	53
X11467A2G	-45	-81	19
X11719G2A	-96	-7	47
X12308A2G	-45	-81	19
X12372G2A	-43	-81	20
X14766C2T	-96	-6	46

Table 1: Mitochondrial DNA Variant - Scaling Dimension Correlations. The variant code is defined by the position on the mitochondrial chromosome, and the transition. For example a transition C2T implies substitution of the reference HG19 base 'C' by base 'T'.

Hierarchical cluster solutions[2] were generated using complete linkage, average linkage and single linkage. Cluster solutions were visually very different indicating that there is no true cluster structure in the data. Figure 2 shows the dendrogram for the single linkage solution.

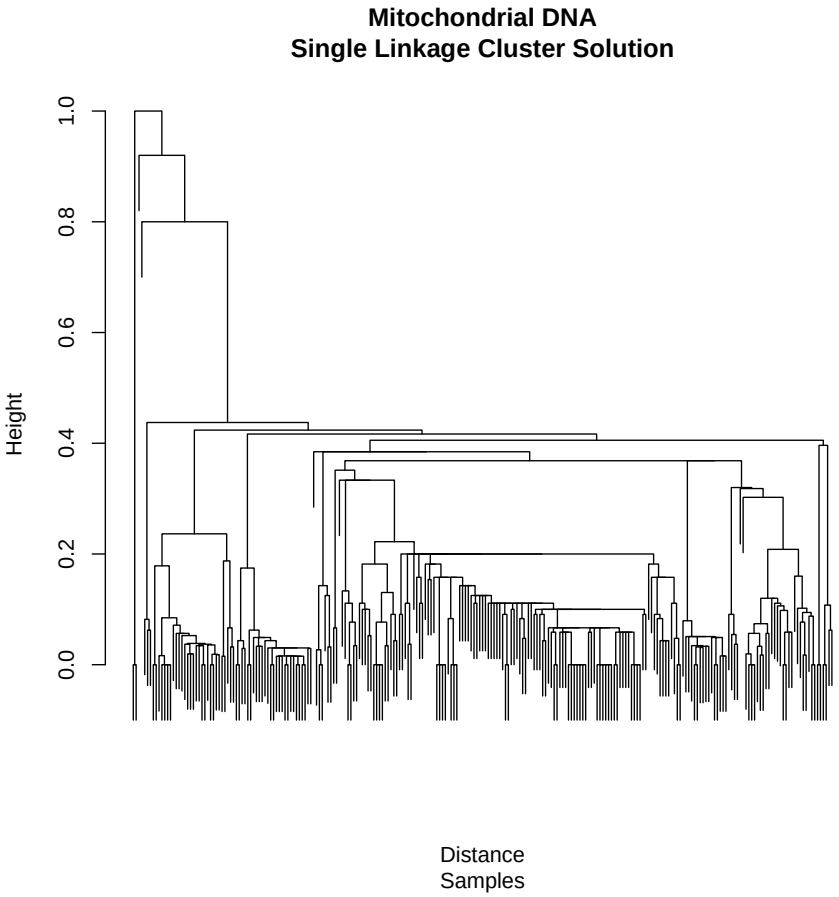


Figure 2: Single Linkage Cluster Solution for Bray Curtis Distance matrix using Variant Incidence.

The strongest feature of the cluster solution is a relatively small number of outlying samples. The cluster tree was cut at a distance value of 0.5: resulting in 8 putative clusters. Membership of clusters is tabulated against exposure group in Table 2. There is no relationship between exposure group and cluster membership.

	Exposure Group:			
Cluster	High	Medium	Low	Control
1	59	32	31	54
2	9	5	2	11
3	2	0	1	3
4	11	4	7	7
5	1	0	0	0
6	0	1	0	2
7	0	1	1	0
8	0	0	1	0

Table 2: Mitochondrial DNA: Cluster Membership by Exposure Group

The structure of the data is dominated by a relatively small number of variants with prevalence between 20% and 80%. There is no indication of between group differences in variant structure.

3 Between Group Differences in Individual Variants

Each variant was analysed by fitting a generalised linear model[4] on exposure group, with a Bernoulli error distribution. A log likelihood χ^2 statistic was calculated.

Variants with a p value ≤ 0.05 are tabulated in Table 3. P values were adjusted using the Benjamini Hochberg procedure[5].

Seven variants showed a statistically significant group effect before correction. Given 224 variants analysed, we expect to find > 10 variants statistically significant by chance alone. After correction to maintain a false discovery rate of 5%, however, none of the variants were statistically significant.

variable	Deviance	Pr(>Chi)	adj.P
X207G2A	9.9	0.019	0.579
X4580G2A	8.2	0.042	0.579
X15904C2T	8.2	0.042	0.579
X10238T2C	7.9	0.049	0.579
X15257G2A	9.3	0.026	0.579
X12007G2A	10.6	0.014	0.579
X16324T2C	9.4	0.024	0.579

Table 3: Chi Square Tests for Variant Between Group Differences

There is no evidence of any difference between groups in the incidence of any of the variants.

References

[1] J. Bray and J. Curtis. An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, 27(4):325–349, 1957.

[2] B. Everitt, Sabine Landau, and Morven Leese. *Cluster Analysis*. Edward Arnold, 4 edition, May 2001.

[3] J. C. Gower. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53(3-4):325–328, 1966.

[4] J. Nelder and R. Wedderburn. Generalised Linear Models. *Journal of the Royal Statistical Society, Series A*, 135:370–384, 1972.

[5] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, 57:289–300, 1995.

Appendix 6
Report from Emphron on Bioinformatic analysis of proteomic data from plasma and leucocytes of exposed workers and controls.

Jet Fuel and Solvents Exposure Project

Analysis of Mass Spectrometry Data

Emphron Informatics Pty Ltd **www.emphron.com**

Report Number:	2012.020
Version Number:	1.0
Client:	Jet Fuel and Solvents Project
Date:	July 22, 2014
Author:	Mervyn Thomas
Email:	mervyn.thomas@emphron.com
Signature:	.....

Executive Summary

1. Data were available from 40 samples, processed using LCMS. One sample was removed because it failed quality control. The remaining 39 samples included 15 control samples and 24 samples from subjects exposed to deseal/reseal solvents.
2. Multivariate exploratory analysis suggested differences between control and exposed subjects in peptide and protein expression. These differences were not large, however.
3. Protein FHOD3 (associated with actin fibre production) appeared to be over expressed in the control subjects. This is of unknown clinical significance.
4. No protein appeared correlated with the health score.
5. A single peptide appeared to be under expressed in exposed subjects, but the mass to charge ratio is consistent with a large number of candidate peptides. This finding is of unknown clinical significance.
6. A single peptide appeared to be negatively correlated with the health score,but the mass to charge ratio is consistent with a large number of candidate peptides. This finding is of unknown clinical significance.

Contents

List of Tables

List of Figures

1 Introduction

1.1 Background and Objectives

The Jet Fuel and Solvents Exposure study has been set up to investigate physiological and other differences between ex Air Force employees who were exposed to aviation fuel during their service. Subjects were stratified into four exposure groups, where Group 4 is a control group (representing no known exposure) and Group 1 is the highest level of exposure.

Data are available from several genomic and proteomic modalities, including: cytogenetics data, High Throughput mRNA gene expression data, micro RNA expression data and from serum analysis by LCMS. Only the results of serum LCMS analysis are described in this report.

Each subject was also asked to complete a health questionnaire, and data from the psycho social scale of the questionnaire were investigated in conjunction with the serum mass spectroscopy.

1.2 Available Data

LCMS data were available from 40 subjects. There were 15 samples from control group subjects (Group 4), 21 samples from high exposure subjects (Group 1), 3 samples from intermediate exposure subjects (Group 2), and 1 sample from the low exposure subjects (Group 3). Since there were very few subjects from low or intermediate exposure groups, subjects were simply coded as control (Group 4) or exposed (Group 1 to 3).

2 Data Processing

2.1 Spectral Averaging

Each sample was subjected to two HPLC runs, and each spotted target from the HPLC run was run through MaldiTOF mass spectroscopy twice: giving a total of four LCMS runs per sample. Peptide and fragment masses within 0.05 Daltons were considered to be duplicates, and the maximum count over each peptide and fragment was taken.

2.2 Protein Estimates

Peptide mz data, and fragment ion mz and signal data were passed to a Mascott search. The highest matching proteins were returned for each sample. Each sample was then represented by a list of proteins and matching Mascot scores.

Proteins that were present in at least 50% of the samples were retained for analysis, and the data was represented as a sample by protein matrix, with each element representing the Mascot score for the given protein in the given sample. A score of zero was imputed when a protein was not identified in a given sample.

Mascot scores are considered semi-quantitative., For a given protein in a matrix of fixed composition, Mascot score will increase monotonically with protein concentration. The relationship, however, is not interval or ratio[9]. Doubling the concentration will not necessarily double the Mascot score. Moreover, it is not necessarily the case that for two different proteins, the protein with the highest Mascot score in a sample will have the highest concentration.

$\log_2$ transformed Mascot scores were analysed, to stabilise variances.

2.3 Peptide Analysis

Peptide mz values for each sample were aggregated across HPLC and MS runs, with a window width of 0.05 Daltons. The maximum signal was chosen for the binned mz value

Peptides that were present in 75% or more of the samples were retained for analysis, and a data matrix was generated with rows defined by samples and columns by peptide mz values. Each cell represented the maximum signal for the given mz value and sample.

3 Exploratory Analysis

3.1 Exploratory Analysis of Proteins

A Bray-Curtis dissimilarity matrix[2] was generated, to characterise the difference between samples. This dissimilarity measure is zero when two samples have the same distribution of Mascot scores across proteins, and attains a maximum value (1) two samples have no proteins with non zero Mascot scores in common.

An image of the between sample dissimilarity matrix is shown in Figure 1. The dissimilarity matrix reveals that one sample, 1030, is markedly different from all of the other samples. Although this sample gave apparently successful HPLC performance, its mass spectroscopy data had previously been identified as marginal. This sample was excluded from further analysis.

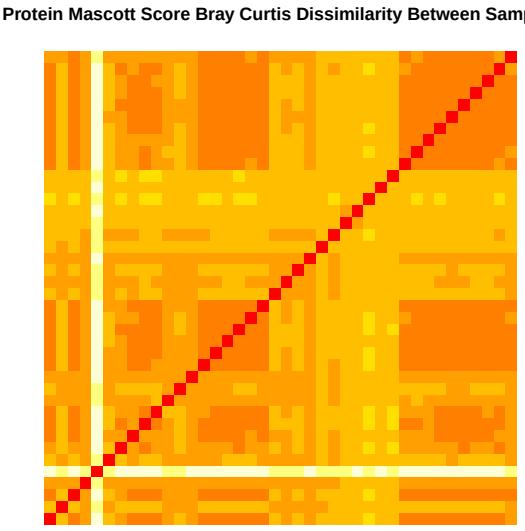


Figure 1: Protein Mascot Score Bray Curtis Dissimilarity Matrix for All Samples

Figure 2 shows the Bray Curtis matrix excluding sample 1030. The rows and columns of the dissimilarity matrix are organised in sample order, with Group 1 at the bottom left, and Group 4 at the top right. Inspection of the matrix images reveals that ten of the control samples (Group 4) have low dissimilarity (the dark block at the upper right of the image), but that these samples are quite dissimilar to the remaining 5 control samples. There is no clear group structure to the dissimilarity matrix.

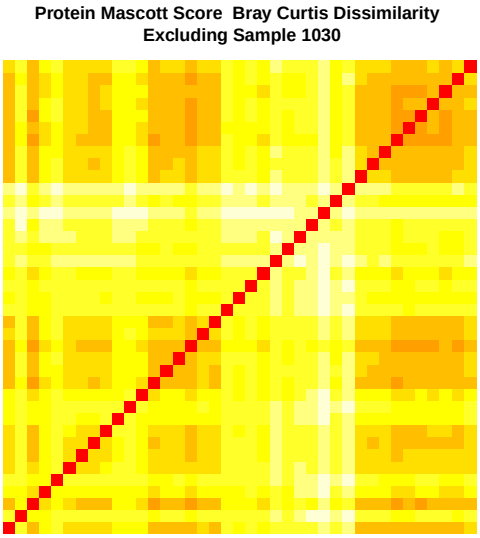


Figure 2: Protein Mascot Score Dissimilarity Matrix Excluding 1030

Single, average and complete linkage cluster dendrograms are shown in Figures 3, 4 and 5 respectively. The average and single linkage dendrograms show chaining[3]. This is a common phenomenon when there is no true cluster structure. The complete linkage solution identifies a group of 5 control chips as being somewhat different from the other chips, but otherwise there is no clear group structure. Given the marked differences between the complete linkage and the other solutions, there is no evidence of any real multivariate structure to the samples.

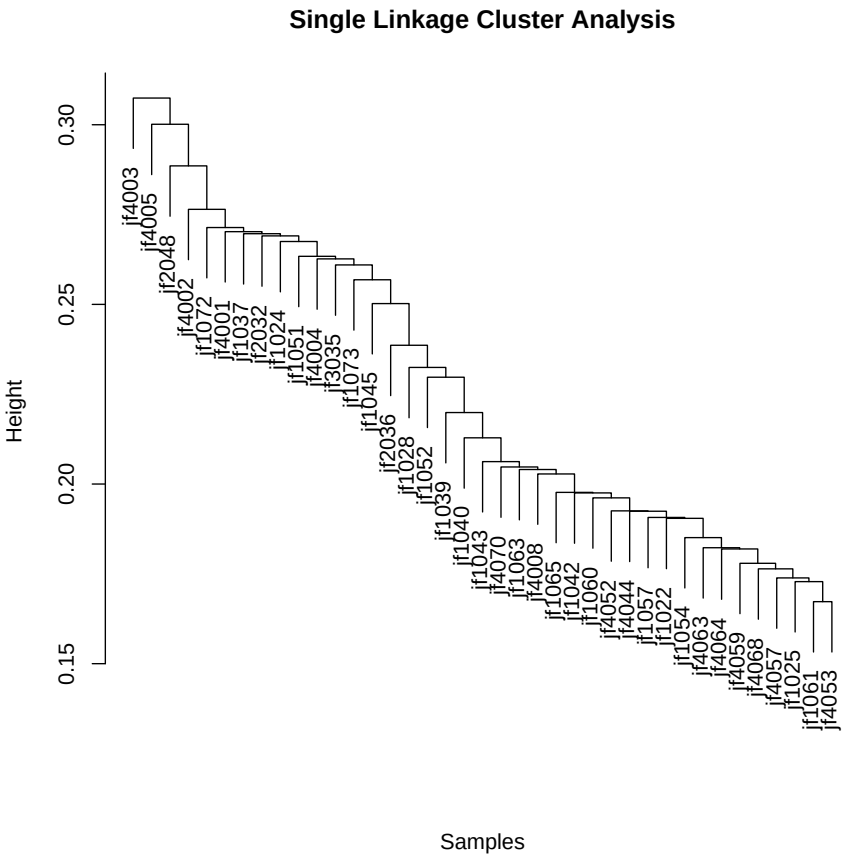


Figure 3: Single Linkage Cluster Analysis For Samples Based on Bray Curtis Dissimilarity for Protein Mascot Scores

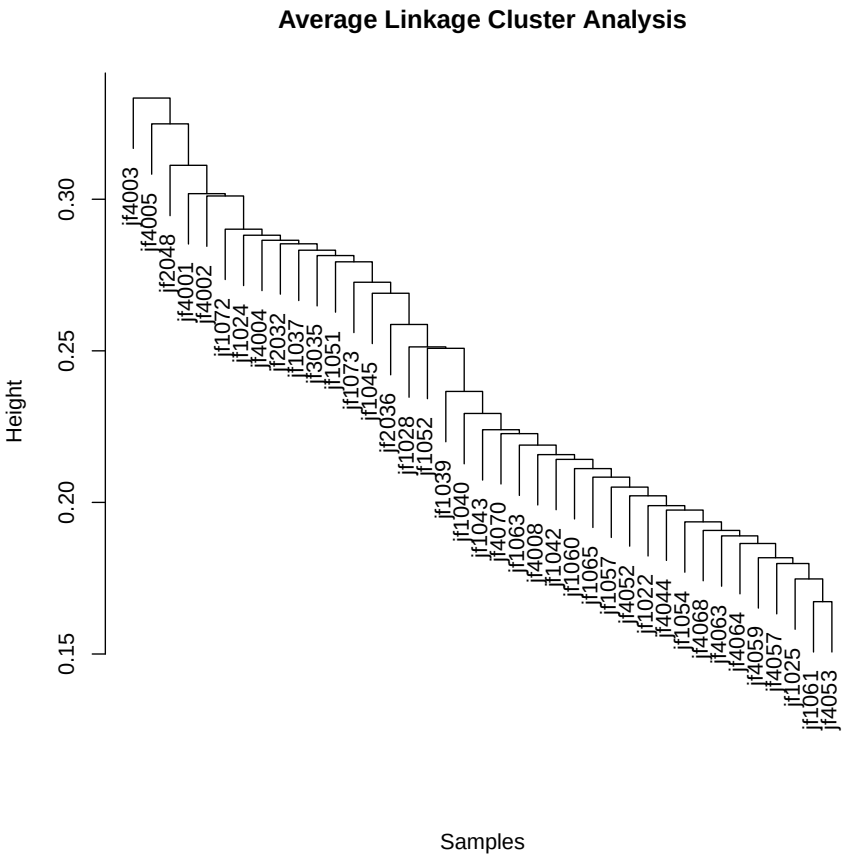


Figure 4: Average Linkage Cluster Analysis For Samples Based on Bray Curtis Dissimilarity for Protein Mascot Scores

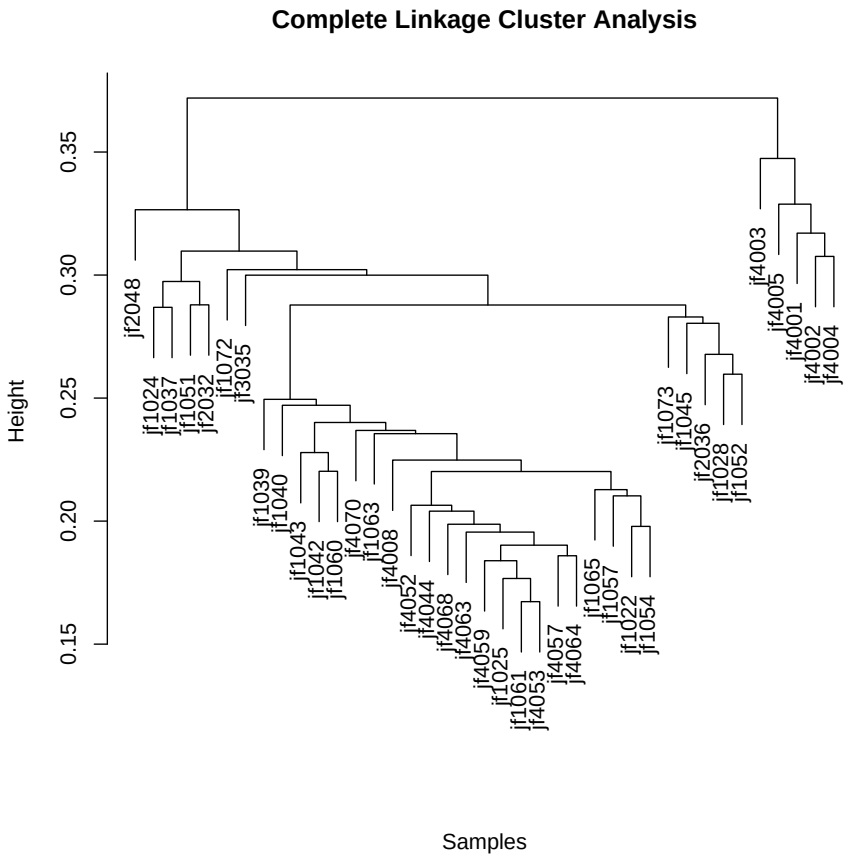


Figure 5: Complete Linkage Cluster Analysis For Samples Based on Bray Curtis Dissimilarity for Protein Mascot Scores

A three dimensional non-metric scaling[1] solution[6] to the dissimilarity matrix generated a stress of 13.6%: which is considered an acceptable solution.

Figure 6 shows samples plotted in the space of the first and second dimensions from the non metric scaling analysis. There is some suggestion of a separation between control and other samples on dimension 2: though this is by no means strong.

[1]Non metric scaling seeks to find a set of coordinates in k dimensional space, such that the Euclidean distance between sample points in that space approximates a given dissimilarity matrix. The scaling is non-metric in that it assumes a monotonic but not necessarily linear, relationship between Euclidean distance in the coordinate k space and dissimilarity.

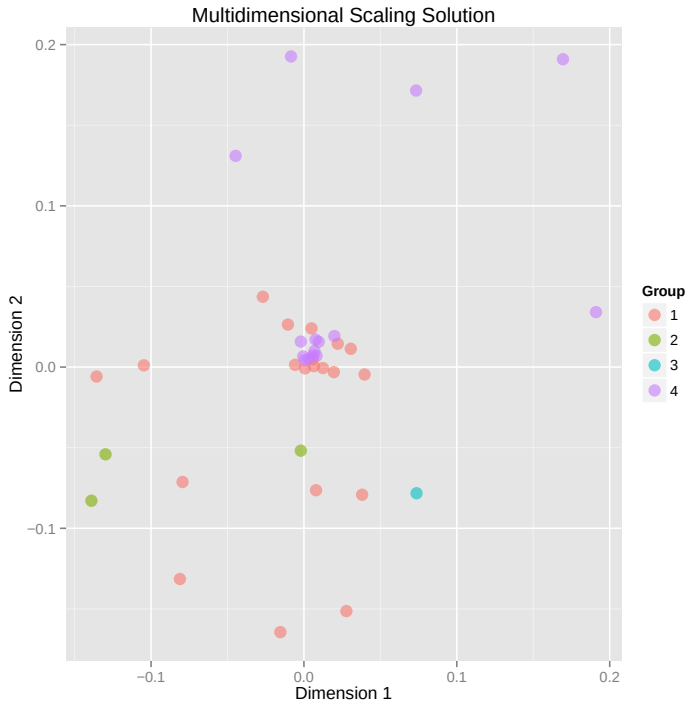


Figure 6: Scatter Plot For First Two Scaling Dimensions From Protein Mascot Score Bray Curtis Dissimilarities

Figure 7 shows box and whisker plots of the three scaling dimensions by exposure group. There is no evidence between group differences for dimensions 1 and 3, but there is evidence of a between groups difference for dimension 2.

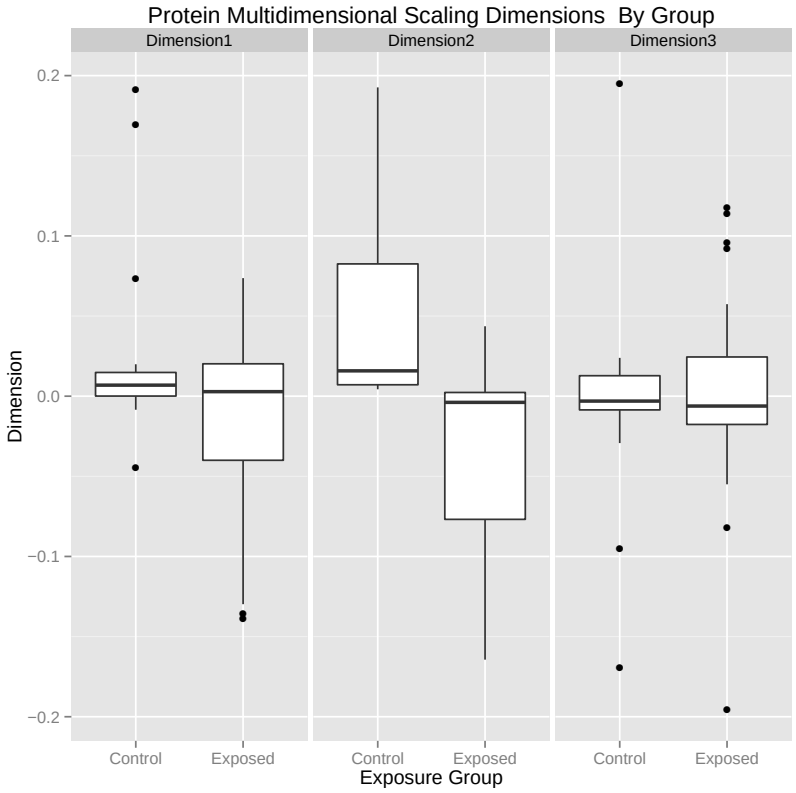


Figure 7: Box And Whisker Plots By Exposure Group For Protein MDS Dimensions from Bray Curtis Dissimilarities.

The analysis of variance for dimension 2 is summarised by Table 1. This conforms the existence of statistically significant between group differences for dimension 2.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Exposure	1	0.07	0.07	17.73	0.0002
Residuals	37	0.16	0.00		

Table 1: Analysis of Variance for Second Scaling Dimension of Protein Mascot Score Bray Curtis Dissimilarity Matrix

3.2 Exploratory Analysis for Peptide Spectra

An image of the between sample peptide dissimilarity matrix is shown in Figure 8. The dissimilarity matrix reveals that one sample, 1030, is markedly different from all of the other samples. Although this sample gave apparently successful HPLC performance, its mass spectroscopy data had previously been identified as marginal. This sample was excluded from further analysis.

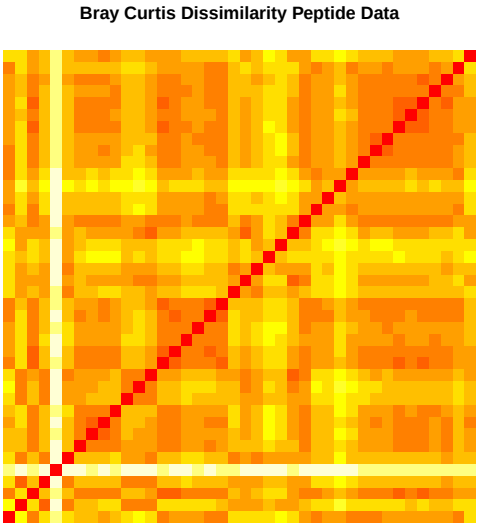


Figure 8: Peptide Bray Curtis Dissimilarity Matrix for All Samples

Figure 9 shows the Bray Curtis matrix excluding sample 1030. The rows and columns of the dissimilarity matrix are organised in sample order, with Group 1 at the bottom left, and Group 4 at the top right. Inspection of the matrix images reveals that nine of the control samples (Group 4) have low dissimilarity (the dark block at the upper right of the image), but that these samples are quite dissimilar to the remaining control samples. There is no clear group structure to the dissimilarity matrix.

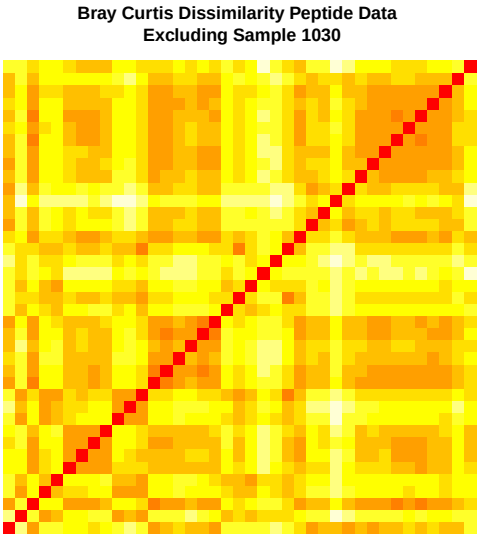


Figure 9: Peptide Bray Curtis Dissimilarity Matrix Excluding 1030

Single, average and complete linkage cluster dendrograms are shown in Figures 10, 11 and 12 respectively.The cluster solutions are visually quite different, indicting that there is no strong cluster solution. There is no suggestion that clusters are associated with exposure group.

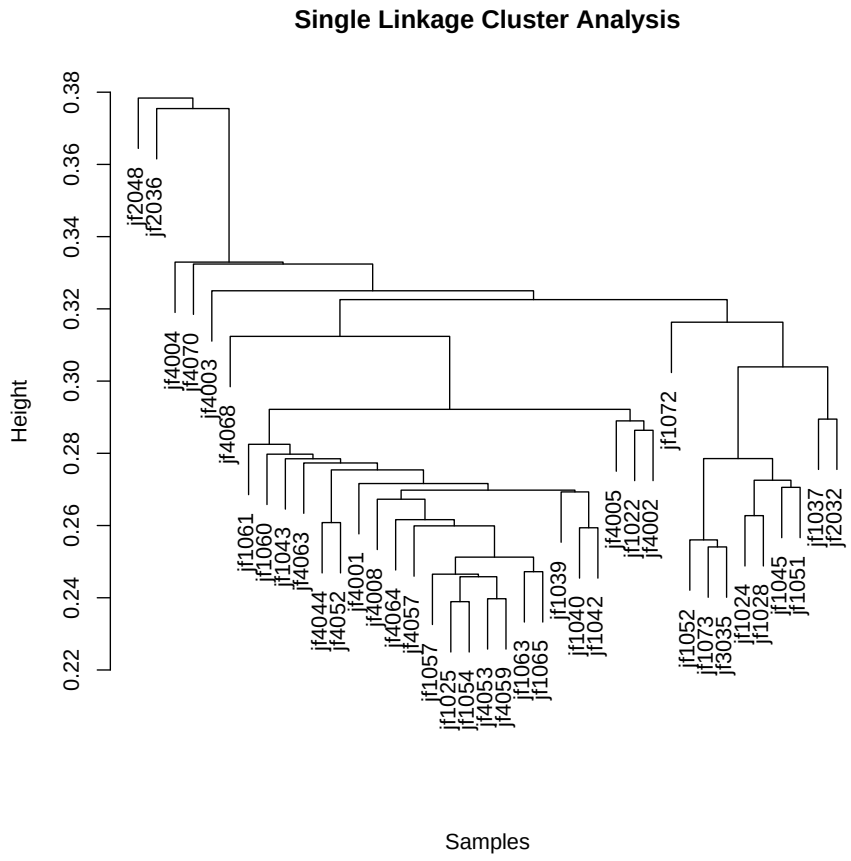


Figure 10: Single Linkage Cluster Analysis For Samples Based on Peptide Bray Curtis Dissimilarities

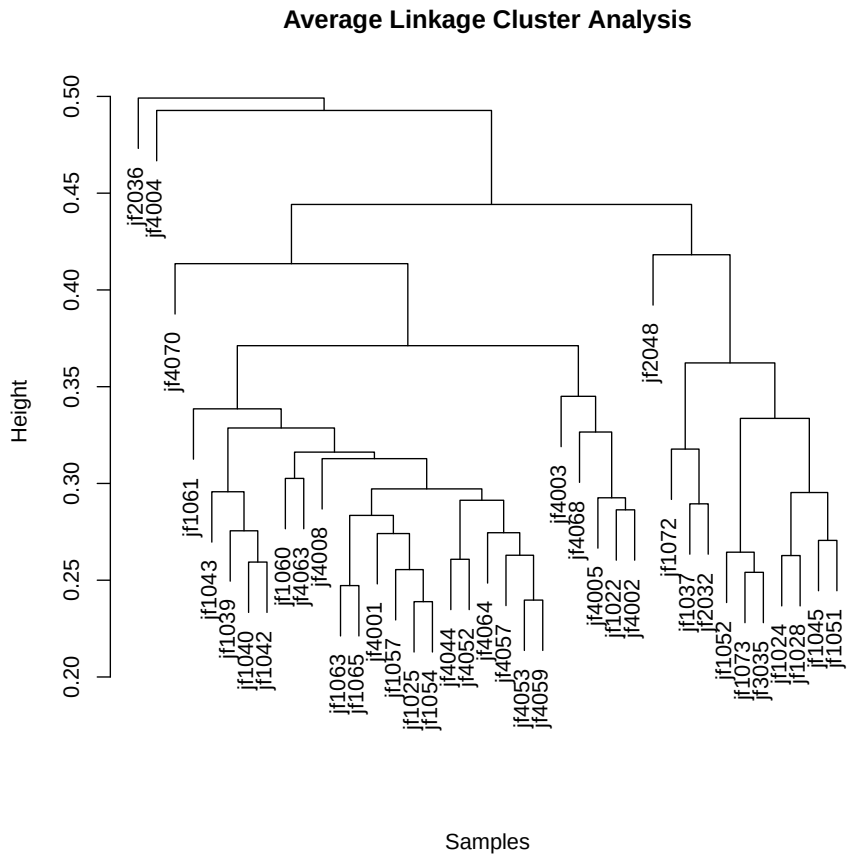


Figure 11: Average Linkage Cluster Analysis For Samples Based on Peptide Bray Curtis Dissimilarities

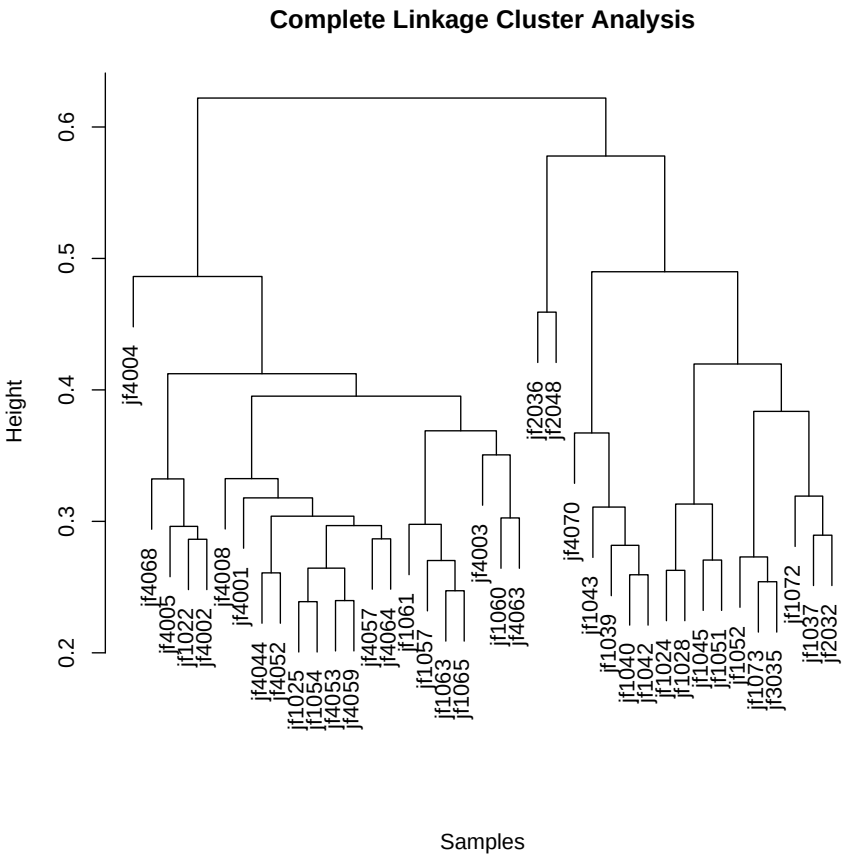


Figure 12: Complete Linkage Cluster Analysis For Samples Based on Peptide Bray Curtis Dissimilarities

A two dimensional non metric scaling solution resulted in a stress of 11 %: which is generally considered an acceptable fit. Figure 13 shows a scatter plot of the two scaling dimensions, with samples coloured according to exposure group. There is no clear evidence of any group separation.

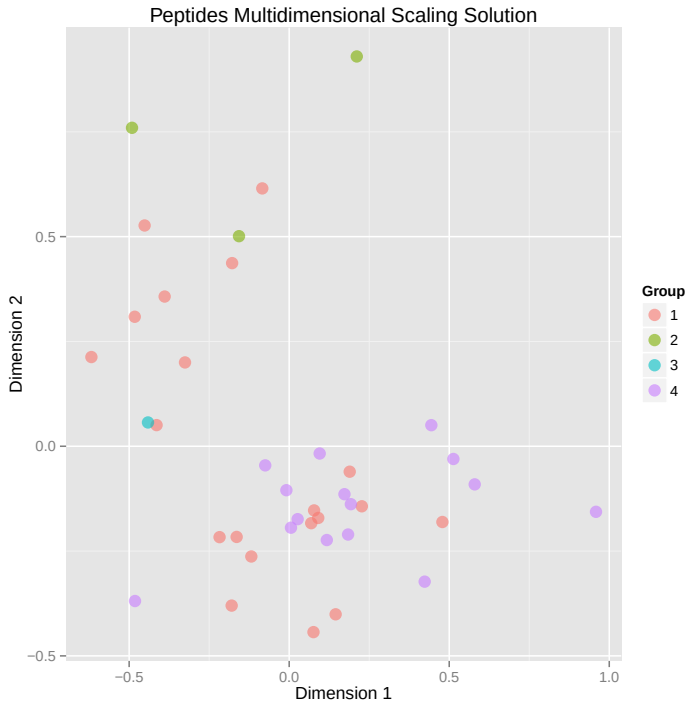


Figure 13: Scatter Plot For Scaling Dimensions From Peptide Bray Curtis Dissimilarities

Figure 14 shows box and whisker plots for the peptide scaling dimensions by exposure group. There is some suggestion that both scaling dimensions differ between groups, though the effects are not large.

Dimension	Effect	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Dimension 1	Exposure	1	1.07	1.07	11.48	0.0017
	Residuals	37	3.46	0.09		
Dimension 2	Exposure	1	0.50	0.50	5.04	0.0308
	Residuals	37	3.65	0.10		

Table 2: Analysis Of Variance For Peptide Scaling Dimensions From Bray Curtis Dissimilarities.

3.3 Summary of Exploratory Analysis Results

- The cluster analysis reveals no suggestion of a strong cluster solution for the peptides or for the proteins. None of the cluster solutions appear to be associated with jet fuel and solvents exposure groups.
- The multidimensional scaling results for the protein dissimilarity suggest a difference between exposure groups on the second scaling dimension, but this difference is not large.

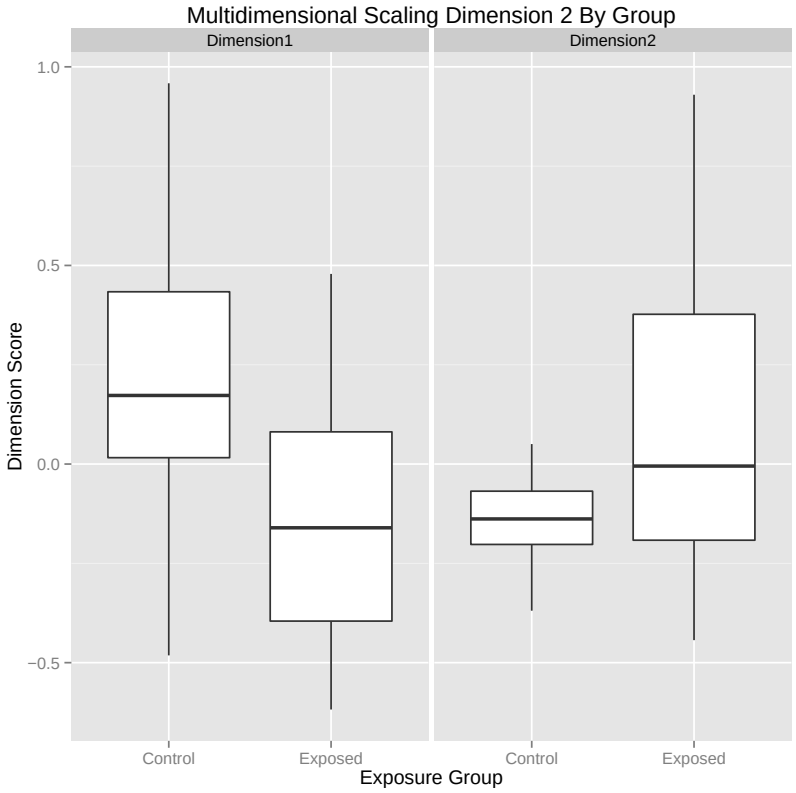


Figure 14: Distribution of Peptide Scaling Dimension Scores from Bray Curtis Dissimilarities By Exposure Group

- The multidimensional scaling results for the peptide dissimilarity suggest a difference between exposure groups on the first and second scaling dimensions, but these differences are not large.

4 Differential Expression

4.1 Methods

Between group differences were evaluated with a linear model, and empirical Bayes moderated t tests[7] based on linear model parameter estimates. These were implemented using Smyth's limma methodology[8]. The between group comparison was between control subjects, and subjects with any levels of exposure (pooled from Groups 1, 2 and 3).

Individual p values for each protein or peptide species were adjusted to maintain a false discovery rate of 5% using the Benjamini Hochberg procedure[10].

This procedure was adopted separately for proteins (using $\log_2$ Mascot scores) and peptides (using $\log_2$ signals).

The empirical Bayes analysis was then repeated using health score as a covariate. Individual p values for the health score covariate for each protein or peptide species were adjusted to maintain a false discovery rate of 5% using the Benjamini Hochberg procedure[10].

This procedure was adopted separately for proteins (using $\log_2$ Mascot scores) and peptides (using $\log_2$ signals).

4.2 Proteins

The only protein to be differentially expressed between exposed and control subjects, following Benjamini Hochberg correction was Swiss-Prot[1] ID FHOD3_HUMAN, with recommended protein name 'FH1/FH2 domain-containing protein 3'. The Mascot score was 16 fold greater in the exposed subjects than in the control subjects. Box and whisker plots are shown in Figure 15.

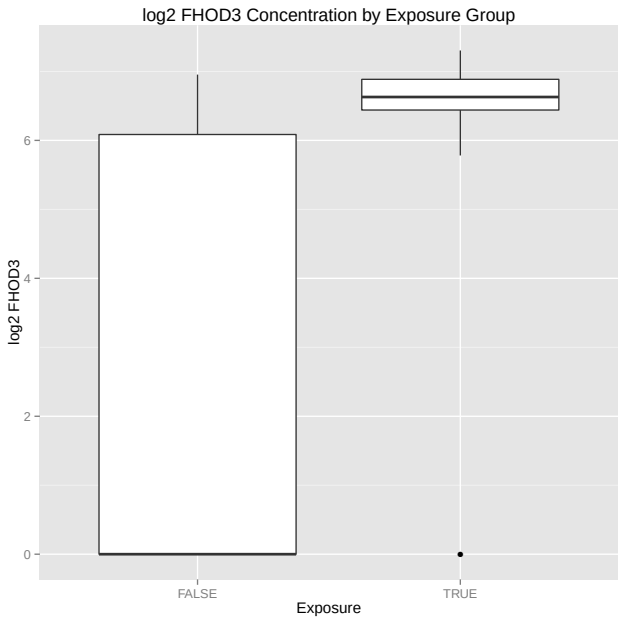


Figure 15: Differential Expression of 'FH1/FH2 domain-containing protein 3' In Control and Exposed Subjects.

FHOD3 is an actin-organising protein that may be associated with stress fibre formation together with cell elongation. It has an essential role in caridogenesis[5]. It appears to be down-regulated in human and murine dilated cardiomyopathy[4].

We have been unable to discover published neurological associations with FHOD3. The messenger RNA associated with this protein was not differentially expressed in the gene expression analysis.

No proteins were statistically significantly associated with the health score following Benjamini Hochberg correction.

4.3 Peptides

Only one peptide showed differential expression between exposed and control subjects following Benjamini Hochberg correction. This peptide had a Mass to charge ratio of 2600.3 and was down regulated (an approximately 5 fold change) in exposed subjects. This peptide is therefore not a constituent of FHOD3 which was up-regulated in exposed subjects. The box and whisker plot for this peptide is shown in Figure 16.

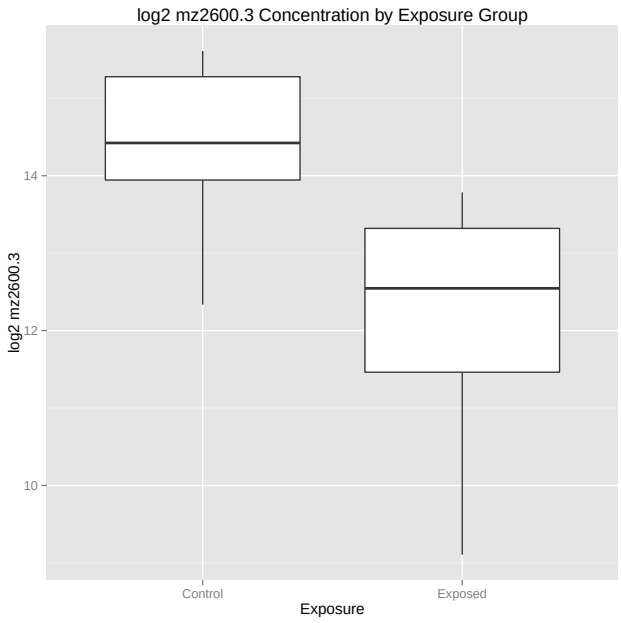


Figure 16: Differential Expression of Peptide With Mass To Charge Ratio 2600.3.

Only one peptide, with mass to charge ratio 3072.4 showed a statistically significant relationship with health score following Benjamini Hochberg correction. This relationship is illustrated in Figure 17. The relationship is strongly influenced by one point which is extreme both for health score and for peptide signal.

4.4 Summary of Differential Expression Results

1. There is evidence that Protein FHOD3 is over expressed in the exposed group. The clinical significance of this finding is uncertain.

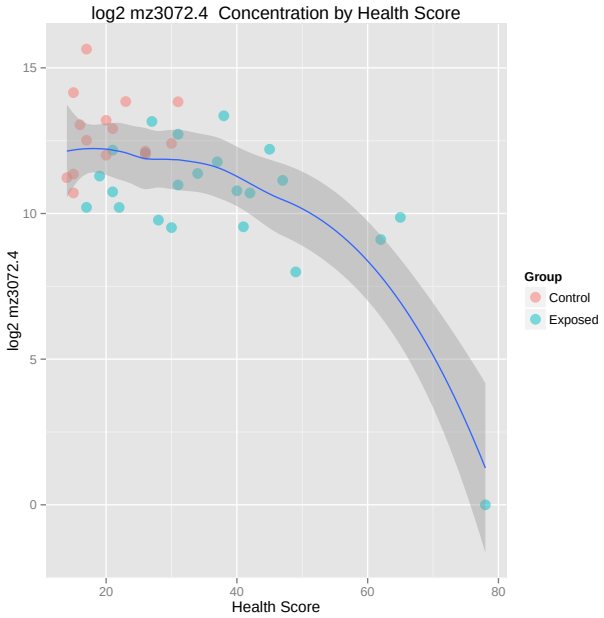


Figure 17: Relationship of Signal For peptide With mass to Charge Ratio 3072.4 with Health Score.

2. No protein appears to be correlated with the health score.
3. A peptide with mass to charge ratio 2600.3 appears to be under expressed in the exposed group. This is of unknown clinical significance. Unfortunately the set of peptides with this mass to charge ratio is too large for useful interpretation.
4. A peptide with mass to charge ratio 3072.4 appears to be negatively correlated with the health score. There are too many peptides with this mass to charge ratio for this result to be clinically useful.

References

[1] E. Boutet, D. Lieberherr, M. Tognolli, M. Schneider, and A. Bairoch. Uniprotkb/swiss-prot. In D. Edwards, editor, *Plant Bioinformatics*, volume 406 of *Methods in Molecular Biology*, pages 89–112. Humana Press, 2007.

[2] J. Bray and J. Curtis. An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, 27(4):325–349, 1957.

[3] B. Everitt, Sabine Landau, and Morven Leese. *Cluster Analysis*. Edward Arnold, 4 edition, May 2001.

[4] T. Iskratsch, S. Lange, J. Dwyer, A. L. Kho, C. d. Remedios, and E. Ehler. Formin follows function: a muscle-specific isoform of FHOD3 is regulated by CK2 phosphorylation and promotes myofibril maintenance. *The Journal of Cell Biology*, 191(6):1159–1172, Dec 2010.

[5] M. Kan-O, R. Takeya, T. Abe, N. Kitajima, M. Nishida, R. Tominaga, H. Kurose, and H. Sumimoto. Mammalian formin Fhod3 plays an essential role in cardiogenesis by organizing myofibrillogenesis. *Biology Open*, 1(9):889–896, Sep 2012.

[6] J. B. Kruskal and M. O. Wish. *Multidimensional Scaling*. Number 07-011 in Sage University Paper series on Quantitative Applications in the Social Sciences. Sage Publications, Newbury Park, CA., 1978.

[7] I. Lonnstedt and T. Speed. Replicated Microarray Data. *Statistica Sinica*, pages 31–46, 2002.

[8] G. K. Smyth. Limma: linear models for microarray data. In R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, and W. Huber, editors, *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, pages 397–420, New York, 2005. Springer.

[9] S. S. Stevens. On the Theory of Scales of Measurement. *Science*, 103(2684):677–680, 1946.

[10] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, 57:289–300, 1995.

Appendix 7
List of abbreviations

ADHREC	Australian Defence Human Research Ethics Committee
AML	acute myeloid leukaemia
ATP	adenosine tri-phosphate
CBC	complete blood count
CHS	commercial human serum
CRF	case report form
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSRS	Deseal/Reseal
DVA	Department of Veterans Affairs
FBC	full blood count
FBS	foetal bovine serum
GHQ 30	General Health Questionnaire 30
IARC	International Agency for Research on Cancer
IOC	International Olympic Committee
ITA	Intent-to-analyse
JFES	Jet Fuel Exposure Study
JIP	Jet fuel infused plasma
JP-8	Jet propulsion 8
MAYAC	Mater Adolescent and Young Adult Clinic
MDS	myelodysplastic syndromes
MHS HREC	Mater Health Services Human Research Ethics Committee
miRNA	micro RNA
MRI(UQ)	Mater Research Institute (University of Queensland)
NHL	non-Hodgkin lymphoma
NHMRC	National Health and Medical Research Council
PP	Per-Protocol
RAAF	Royal Australian Air Force
RNA	Ribonucleic acid
RSL	Returned Services League
SAP	Statistical Analysis Plan
SD	Standard deviation
SHOAMP	Study of Health Outcomes in Aircraft Maintenance Personnel
SNP	Single nucleotide polymorphism
WBC	White blood cells

Appendix 8 Glossary

Apoptosis

is the process of programmed cell death that may occur in multicellular organisms

SNP array

is a type of DNA microarray which is used to detect polymorphisms within a population, it is a useful tool for studying slight variations between whole genomes

Bioinformatics

combines computer science, statistics, mathematics and engineering to study and process biological data

Controls

a group of people who have had no exposure to jet fuel. Controls are used to increase the reliability of results.

Cytotoxicity

is the quality of being toxic to cells.

Epigenetics

is the study of changes in gene expression caused by certain base pairs in DNA, or RNA, being “turned off” or “turned on” again, through chemical reactions.

Gene transcription/translation

cells use a two-step process of transcription and translation to read each gene and produce the string of amino acids that make up a protein.

Genomes

the complete set of DNA within a single cell of an organism.

Genomics

is a discipline in genetics to sequence, assemble, and analyse the function and structure of genomes

Mass spectrometry

is an analytical chemistry technique that measures the mass-to-charge ratio and abundance of gas-phase ions.

Methylation

denotes the addition of a methyl group to a substrate or the substitution of an atom or group by a methyl group.

Mitochondria

are membrane-bound organelles found in most animal and plant cells.

Proteomics

is the large-scale study of proteins, particularly their structures and functions.

workers

Australian Air Force personnel working on the Deseal/Reseal project.

Appendix 9
References

Adams JM, Cory S. Bcl-2-regulated apoptosis: mechanism and therapeutic potential. *Curr Opin Immunol* **19**, 488-496 (2007).

Adams JM, Cory S. The Bcl-2 apoptotic switch in cancer development and therapy. *Oncogene* **26**, 1324-1337 (2007).

Australian Institute of Health and Welfare. Third Study of Mortality and Cancer Incidence in Aircraft Maintenance Personnel: A Continuing Study of F1-11 Deseal/Reseal Personnel. 2009.

Bayat A. Science, medicine, and the future: bioinformatics. BMJ 324 (2002) 1018–1022

Bleier J, McFarlane A, McGuire A, Treloar S, Waller M, Dobson A. Risk of adverse health outcomes associated with frequency and duration of deployment with the Australian Defence Force. Mil Med. 2011 Feb;176(2): 139-46.

Bollati V, Baccarelli A, Hou L, et al. Changes in DNA methylation patterns in subjects exposed to low-dose benzene. Cancer Res. 67 (2007) 876–880.

Bollati V, Baccarelli A. Environmental Epigenetics (review). Heredity. 2010, February 24 [advance online publication]

Briefing notes to Chief of Air Force on research into Mitochondrial Disease in DeSeal/ReSeal personnel, 23 January 2008.

Briefing notes to the Commonwealth of Australia Parliamentary Committee Hearing, 16 April 2009.

Brittion DA, Okoniewski MJ, Connolly Y, Miller CJ. Exon level integration of proteomics and microarray data. BMC Bioninformatics. 2008;9:118.

Cho JY. Suppressive effect of hydroquinone, a benzene metabolite, on in vitro inflammatory responses mediated by macrophages, monocytes and lymphocytes. Mediat. Inflamm. (2008) 298010.

Cline MS, Smoot M, Cerami E, et al. Integration of biological networks and gene expression data using Cytoscape. Nat. Protoc. 2 (2007) 2366–2382.

Crowley JJ, Sullivan PF, McLeod HL. Pharmacogenomic genome-wide association studies: lessons learned thus far. Pharmacogenomics 10 (2009) 161–163.

Dixon-McIver A, East P, Mein CA, et al. Distinctive patterns of microRNA expression associated with karyotype in acute myeloid leukaemia, PLoS ONE 3 (2008) e2141

Dolinoy DC, Jirtle RL. Environmental epigenomics in human health and disease. Environ Mol Mutagen. 49 (2008) 4–8.

Edwards SW, Preston RJ. Systems biology and mode of action based risk assessment. Toxicol Sci. 106 (2008) 312–318.

Ellis L, Atadja PW, Johnstone RW. Epigenetics in cancer: targeting chromatin modifications. *Mol Cancer Ther* **8**, 1409-1420 (2009).

Forrest MS, Lan Q, Hubbard AE, et al. Discovery of novel biomarkers by microarray analysis of peripheral blood mononuclear cell gene expression in benzene-exposed workers. Environ. Health Perspect. 113 (2005) 801– 807.

Frew AJ, Johnstone RW, Bolden JE. Enhancing the apoptotic and therapeutic effects of HDAC inhibitors. *Cancer Lett* **280**, 125-133 (2009).

Galvan N, Lim S, Zmugg S, Smith MT, Zhang L. Depletion of WRN enhances DNA damage in HeLa cells exposed to the benzene metabolite hydroquinone. Mutat. Res. 649 (2008) 54–61.

Glass DC, Gray CN, Jolley DJ, et al. Leukemia risk associated with low-level benzene exposure. Epidemiology (Cambridge, MA) 14 (2003) 569–577.

Goldstein JC, Waterhouse NJ, Juin P, Evan GI, Green DR. The coordinate release of cytochrome c during apoptosis is rapid, complete and kinetically invariant. *Nat Cell Biol* **2**, 156-162 (2000).

Hart D, Venter D. Dendritic cells in leukemia and lymphoma. *Leuk Lymphoma*. **47**(4):613-22.

Hayes RB, Songnian Dosemeci YM, Linet M. Benzene and lymphohematopoietic malignancies in humans. Am. J. Ind. Med. 40 (2001) 117–126.

Hoboken, NJ, Smith M, Rappaport S. Building exposure biology centres to put the E into “G°øE” interaction studies. Environ. Health Perspect. 117 (2009) A334–A335.

International Agency for Research on Cancer (IARC). Some Industrial Chemicals and Dyestuffs. Monogr. Eval. Carcinog. Risks Hum., 1982.

International Agency for Research on Cancer (IARC). Monographs on the Evaluation of Carcinogenic Risks to Humans Overall Evaluations of Carcinogenicity, An Updating of IARC Monographs. 1987.

Ji Z, Zhang L, Guo W, McHale CM, Smith MT. The benzene metabolite, hydroquinone and etoposide both induce endoreduplication in human lymphoblastoid TK6 cells. Mutagenesis 24 (2009) 367–372.

Jirtle RL, Skinner MK. Environmental epigenomics and disease susceptibility. Nat. Rev. 8 (2007) 253–262.

Kim E, Kang BY, Kim TS. Inhibition of interleukin-12 production in mouse macrophages by hydroquinone, a reactive metabolite of benzene, via suppression of nuclear factor-kappaB binding activity. Immunol. Lett. 99 (2005) 24–29.

Kirschner MW. The meaning of systems biology. Cell 121 (2005) 503–504.

Kitano H. Systems biology: a brief overview. Science (New York, NY) 295 (2002) 1662–1664.

Knight JA, Skol AD, Shinde A, et al. Genome-wide association study to identify novel loci associated with therapy-related myeloid leukemia susceptibility. Blood 113 (2009) 5575–5582.

Lan Q, Zhang L, Li G, et al. Hematotoxicity in workers exposed to low levels of benzene. Science (New York, NY) 306 (2004) 1774–1776.

Lan Q, Vermeulen R, Zhang L, et al. Benzene exposure and hematotoxicity. Science (New York, NY) 312 (2006) 998–999.

Lan Q, Zhang L, Shen M, et al. Large-scale evaluation of candidate genes identifies associations between DNA repair and genomic maintenance and development of benzene hematotoxicity. Carcinogenesis 30 (2009) 50–58.

Lan Q, Zhang L, Shen M, et al. Polymorphisms in cytokine and cellular adhesion molecule genes and susceptibility to hematotoxicity among workers exposed to benzene. Cancer Res. 65 (2005) 9574–9581.

Lebaron MJ, Rasoulpour RJ, Klapacz J, Ellis-Hutchings RG, Hollnagel HM, Gollapudi BB. Epigenetics and chemical safety assessment. Mutat Res. 2010, April 23 [Epub ahead of print]

Liao TT1, Jia RW, Shi YL, Jia JW, Wang L, Chua H. Propidium iodide staining method for testing the cytotoxicity of 2,4,6-trichlorophenol and perfluorooctane sulfonate at low concentrations with Vero cells. J Environ Sci Health A Tox Hazard Subst Environ Eng. 2011;46(14):1769-75.

Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics (Oxford, England) 21 (2005) 3448–3449.

Martin SJ*, et al.* Early redistribution of plasma membrane phosphatidylserine is a general feature of apoptosis regardless of the initiating stimulus: inhibition by overexpression of Bcl-2 and Abl. *J Exp Med* **182**, 1545-1556 (1995).

Mattingly CJ, Rosenstein MC, Colby GT, Forrest Jr. JN, Boyer JL. The Comparative Toxicogenomics Database (CTD): a resource for comparative toxicological studies. J. Exp. Zool. A: Comp. Exp. Biol. 305 (2006) 689–692.
McHale C, Zhang L, Hubbard AE, Smith M. Toxicogenomic studies in human populations, in: Boverhof DR, Gollapudi BB (Eds.), Applications of Toxicogenomics in Safety Evaluation and Risk Assessment, John Wiley & Sons, Inc., p177-206.

McHale CM, Zhang L, Lan Q, et al. Changes in the peripheral blood transcriptome associated with occupational benzene exposure identified by cross-comparison on two microarray platforms. Genomics 93 (2009) 343–349.

McHale C, Zhang L, Lan Q, et al. Low-dose, occupational exposure to the leukemogen benzene induces robust changes in the blood transcriptome associated with altered immune system biology. American Society of Hematology, San Francisco, CA, 2008.

Mota NP, Medved M, Whitney D, Hiebert-Murphy D, Sareen J. Protective factors for mental disorders and psychological distress in female, compared with male, service members in a representative sample. Can J Psychiatry. 2013 Oct;58(10):570-8.

Muller CI, Luong QT, Shih LY, et al. Identification of marker genes including RUNX3 (AML2) that discriminate between different myeloproliferative neoplasms and normal individuals. Leukemia 22 (2008) 1773–1778.

National Research Council (NRC), Toxicity Testing in the 21st Century: A Vision and a Strategy, Committee on Toxicity Testing and Assessment of Environmental Agents, National Academies Press, Washington, DC, 2007.

Neufeld G, Shraga-Heled N, Lange T, et al. Semaphorins in cancer. Front. Biosci. 10 (2005) 751–760.

North M, Shuga J, Loguinov A, et al. Utilizing functional genomics in yeast to discover novel biomarkers of benzene toxicity in humans, in: International Symposium on Benzene Health Effects and Mechanisms of Bone Marrow Toxicity, Munich, Germany, 2009.

Ouyang Y, Virasch N, Hao P, et al. Suppression of human IL-1beta, IL-2, IFNgamma, and TNF-alpha production by cigarette smoke extracts. J. Allergy Clin. Immunol. 106 (2000) 280–287.

Paschka P. Core binding factor acute myeloid leukemia. Semin. Oncol. 35 (2008) 410–417.

Pedersen-Bjergaard J, Andersen MK, Andersen MT, Christiansen DH. Genetics of therapy-related myelodysplasia and acute myeloid leukemia. Leukemia 22 (2008) 240–248.

Ren X, Lim S, Smith MT, Zhang L. Werner syndrome protein, WRN, protects cells from DNA damage induced by the benzene metabolite hydroquinone. Toxicol. Sci. 107 (2009) 367–375.

Ross D, Siegel D, Schattenberg DG, Sun XM, Moran JL. Cell-specific activation and detoxification of benzene metabolites in mouse and human bone marrow: identification of target cells and a potential role for modulation of apoptosis in benzene toxicity. Environ. Health Perspect. 104 (Suppl. 6) (1996) 1177–1182.

Savitz DA, Andrews KW. Review of epidemiologic evidence on benzene and lymphatic and hematopoietic cancers. Am. J. Ind. Med. 31 (1997) 287–295.

Sedelies KA*, et al.* Blocking granule-mediated death by primary human NK cells requires both protection of mitochondria and inhibition of caspase activity. Cell Death Differ **15**, 708-717 (2008).

Shen M, Zhang L, Bonner MR, Liu CS, Li G, Vermeulen R, Dosemeci M, Yin S, Lan Q. Association between mitochondrial DNA copy number, blood cell counts, and occupational benzene exposure. Environ Mol Mutagen **49**(6): 453-7. (2008)

Shen M, Lan Q, Zhang L, et al. Polymorphisms in genes involved in DNA double-strand break repair pathway and susceptibility to benzene-induced hematotoxicity. Carcinogenesis 27 (2006) 2083–2089

Singh DK, Ahn B, Bohr VA. Roles of RECQ helicases in recombination based DNA repair, genomic stability and aging. Biogerontology 10 (2009) 235– 252.

Skibola CF, Bracci PM, Halperin E, et al. Genetic variants at 6p21.33 are associated with susceptibility to follicular lymphoma. Nat. Genet. 41 (2009) 873–875.

Slade T, Grove R, Burgess P, Kessler RC. Psychological Distress Scale: normative data from the 2007 Australian National Survey of Mental Health and Wellbeing. Aust N Z J Psychiatry. 2011 Apr;45(4):308-16.

Smith MT, Jones RM, Smith AH. Benzene exposure and risk of non-Hodgkin lymphoma. Cancer Epidemiol. Biomark. Prev. 16 (2007) 385–391.

Smith MT. The mechanism of benzene-induced leukemia: a hypothesis and speculations on the causes of leukemia. Environ. Health Perspect. 104 (Suppl. 6) (1996) 1219–1225.

Smith MT, Zhang L, Wang Y, et al. Increased translocations and aneusomy in chromosomes 8 and 21 among workers exposed to benzene. Cancer Res. 58 (1998) 2176–2181.

Snyder R, Hedli CC, An overview of benzene metabolism. Environ. Health Perspect. 104 (Suppl. 6) (1996) 1165–1171.

Study of Health Outcomes in Aircraft Maintenance Personnel (SHOAMP), Phase II Mortality and Cancer Incidence Study Interim Report, July 2003, Australian Department of Defence, Canberra; July 2003.

Subrahmanyam VV, Ross D, Eastmond DA, Smith MT. Potential role of free radicals in benzene-induced myelotoxicity and leukemia. Free Radic Biol Med. 11 (1991) 495–515.

Trosko JE. Commentary on "Toxicity testing in the 21st century: a vision and a strategy": stem cells and cell-cell communication as fundamental targets in assessing the potential toxicity of chemicals. Hum Exp Toxicol. 2010; 29(1): 21-29.

Van der Laan MJ, Pollard KS. A new algorithm for hybrid hierarchial clustering with visualization and the bootstrap. J Stat Plann Inference (2003) 275–303.

Vermeulen R, Lan Q, Zhang L, et al. Decreased levels of CXC-chemokines in serum of benzene-exposed workers identified by array-based proteomics. Proc Natl Acad Sci. U.S.A. 102 (2005) 17041–17046.

Vlaanderen J, Moore LE, Smith MT, et al. Application of omics technologies in occupational and environmental health research: current status and projections. Occup Environ Med. (2009).

Warrener R*, et al.* Tumor cell-selective cytotoxicity by targeting cell cycle checkpoints. Faseb *J* **17**, 1550-1552 (2003).

Waterhouse NJ, Goldstein JC, Kluck RM, Newmeyer DD, Green DR. The (Holey) study of mitochondria in apoptosis. Methods Cell Biol 66, 365-391 (2001).

Waterhouse NJ, Trapani JA. A new quantitative assay for cytochrome c release in apoptotic cells. Cell Death Differ **10**, 853-855 (2003).

Waterhouse NJ. The cellular energy crisis: mitochondria and cell death. Med Sci Sports Exerc **35**, 105-110 (2003).

Waterhouse NJ, Ricci JE, Green DR. And all of a sudden it's over: mitochondrial outer-membrane permeabilization in apoptosis. Biochimie **84**, 113-121 (2002).

Waterhouse NJ*, et al.* Functional dissociation of DeltaPsim and cytochrome c release defines the contribution of mitochondria upstream of caspase activation during granzyme B-induced apoptosis. Cell Death Differ **13**, 607-618 (2006).

Waters M, Stasiewicz S, Merrick BA, et al. CEBS—Chemical Effects in Biological Systems: a public data repository integrating study design and toxicity data with microarray and proteomics data. Nucleic Acids Res. 36 (2008) D892– 900.

Witzmann FA, Carpenter RL, Ritchie GD, Wilson CL, Nordholm AF, Ross J. Toxicity of chemical mixtures: Proteomic analysis of persisting liver and kidney protein alterations inducted by repeated exposure of rats to JP-8 jet fuel vapor. Electrophoresis. 2000;21:2138-2147.

Yoon BI, Hirabayashi Y, Kawasaki Y, et al. Aryl hydrocarbon receptor mediates benzene-induced hematotoxicity. Toxicol. Sci. 70 (2002) 150–156.

Zhang L, McHale CM, Rothman N, et al. Systems Biology of human benzene exposure. Chem Biol Interact. 2010;184(1-2):86-93.

Zhang L, Smith M, Bandy B, Tamaki SJ, Davison AJ. Role of Quinones, active oxygen species and metals in the genotoxicity of 1,2,4-benzenetriol, ametabolite of benzene, in: H. Nhol. H. Esterbauer, C. Rice-Evans (Eds.), Free Radicals in the Environment, Medicine and Toxicology, Richelieu, London, 1994, pp. 521–562.

Zhang L, Eastmond DA, Smith MT. The nature of chromosomal aberrations detected in humans exposed to benzene. Crit. Rev. Toxicol. 32 (2002) 1–42.

Zhang L, Rothman N, Wang Y, et al. Increased aneusomy and long arm deletion of chromosomes 5 and 7 in the lymphocytes of Chinese workers exposed to benzene. Carcinogenesis 19 (1998) 1955–1961.