
Federal Aviation
Administration

DOT/FAA/AM-08/5
Office of Aerospace Medicine
Washington, DC 20591

# Functional Genomics Group—Program Description

Dennis Burian

Civil Aerospace Medical Institute
Oklahoma City, OK 73125

February 2008

Final Report

# NOTICE



———————

**Technical Report Documentation Page**

| 1. Report No. | 2. Government Accession No. | 3. Recipient's Catalog No. |
|---|---|---|
| DOT/FAA/AM-08/5 | | |

| 4. Title and Subtitle | 5. Report Date |
|---|---|
| Functional Genomics Group-Program Description | February 2008 |
| | 6. Performing Organization Code |

| 7. Author(s) | 8. Performing Organization Report No. |
|---|---|
| Burian, D | |

| 9. Performing Organization Name and Address | 10. Work Unit No. (TRAIS) |
|---|---|
| FAA Civil Aerospace Medical Institute<br>P.O. Box 25082<br>Oklahoma City, OK 73125 | 11. Contract or Grant No. |

| 12. Sponsoring Agency name and Address | 13. Type of Report and Period Covered |
|---|---|
| Office of Aerospace Medicine<br>Federal Aviation Administration<br>800 Independence Ave., S.W.<br>Washington, DC 20591 | 14. Sponsoring Agency Code |

15. Supplemental Notes

Work was accomplished under approved task AM-

16. Abstract

Regulation of gene expression is a complex process that exquisitely responds to the environment to maintain cellular and, ultimately, organismal homeostasis. Gene expression research is undertaken at the Federal Aviation Administration as a means of discovering sets of biomarkers that change in response to environmental factors that affect aviation safety. This article reviews mechanisms of gene regulation and discusses how genomics is changing the way medicine is practiced today as a means of demonstrating that molecular medicine is here to stay. Next, the protocols that have been developed into a cohesive workflow to perform gene expression analysis are presented. Environmental factors currently under investigation are delineated followed by a discussion of other factors of interest for future research. We believe this research will benefit the aviation industry by improving the accuracy of the data used to write regulation, thus improving the already remarkable safety record of the aviation industry and decreasing medical risks to flight crew.

| 17. Key Words | 18. Distribution Statement |
|---|---|
| Gene Expression; Gene Regulation; Molecular Medicine; Aerospace Medicine; Microarray; Messenger RNA | Document is available to the public through the Defense Technical Information Center, Ft. Belvior, VA 22060; and the National Technical Information Service, Springfield, VA 22161 |

| 19. Security Classif. (of this report) | 20. Security Classif. (of this page) | 21. No. of Pages | 22. Price |
|---|---|---|---|
| Unclassified | Unclassified | 15 | |

**Form DOT F 1700.7** (8-72) Reproduction of completed page authorized

# Contents


INTRODUCTION . . . . . 1
- Classical Modes of Inheritance . . . . . 1
- Non-classical Modes of Inheritance . . . . . 2
- Staying Ahead of the Curve . . . . . 2
- Marker Discovery Projects . . . . . 3

GENE EXPRESSION AT THE FAA . . . . . 4
- Workflow . . . . . 4
- Array Design . . . . . 4
- Analysis . . . . . 6

CURRENT GENE EXPRESSION STUDIES . . . . . 7
- Moderate Alcohol Use . . . . . 7
- Fatigue . . . . . 8
- Radiation . . . . . 8
- Hypoxia/Altitude Exposure . . . . . 8

GENOTYPING STUDIES . . . . . 8

FUTURE STUDIES . . . . . 8

CONCLUSIONS . . . . . 9

REFERENCES . . . . . 10

# FUNCTIONAL GENOMICS GROUP-PROGRAM DESCRIPTION

## INTRODUCTION

Structurally, DNA is a double stranded polymer of molecules called deoxy-nucleotides. Deoxynucleotides contain two chemical moieties, the base and a deoxy-ribose sugar that is modified with a phosphate group. Four nucleotides are used in the synthesis of DNA delineated by the identity of the base, adenine (A), thymidine (T), cytidine (C) or guanidine (G). The polymer forms by the enzymatic addition of the phosphate group from a free nucleotide to one of the hydroxyl groups of the terminal deoxy-ribose of the nascent molecule. During nucleic acid synthesis, the synthesis machinery "reads" the sequence of bases on the opposite strand. In this way, the synthesis machinery is directed to add the complementary base to the newly formed nucleic acid strand, preserving the sequence information present in the original molecule. A and T are complementary, as are C and G.

### Classical Modes of Inheritance

The "classical" Central Dogma of Biology holds that DNA is a storage molecule for genes. In the nucleus, genes are transcribed into RNA that is processed and exported to the cytoplasm where ribosomes translate the messenger RNA (mRNA) into a single protein. Also known as the "One Gene, One Protein" theory, the Central Dogma, as originally hypothesized, held that each gene led to one and only one protein (Fig. 1A). However, the Central Dogma has undergone major revision in recent years (Fig. 1B) as understanding of the many mechanisms that regulate gene expression has expanded. There is an increasing awareness that the human genome encodes far fewer genes than originally thought; current estimates range between 20,000 and 25,000 genes in contrast to estimates that were as high as 100,000 genes ten years ago. The discrepancy is due, in part, to alternative splicing of transcribed RNA resulting in multiple gene products from a single gene. Alternative splicing was the first departure from the Central Dogma to be discovered.

After a gene is transcribed from genomic DNA to RNA, it is spliced to form the messenger RNA (mRNA) sequence. During the course of splicing, pieces of the original transcript are enzymatically removed by a specialized set of proteins, and the pieces that eventually will be translated into protein are rejoined to make the mature mRNA. The pieces that are kept are known as exons; the pieces that are discarded, or intervening sequence, are called introns. Splicing is sequence dependent and tightly controlled. Alternative splicing events arise when certain exons in the gene of interest are either kept or lost depending on environmental factors or cell type. The proteins that result from alternative splicing do not have the same sequence and will have altered activities compared to the same protein that is not alternatively spliced. It should be noted that because alternative splicing can be cell type specific, the definition of "normal" becomes blurred; the mRNA in one cell type would be an aberration in another cell type. Further complicating the picture is the recent discovery that gene duplications have occurred in human evolution such that certain ethnic groups have more copies of some genes than other ethnicities (1).

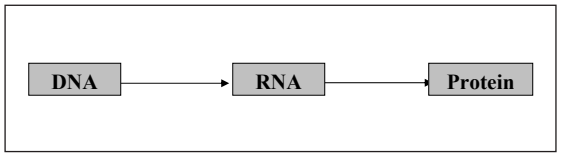


**Figure 1A.**

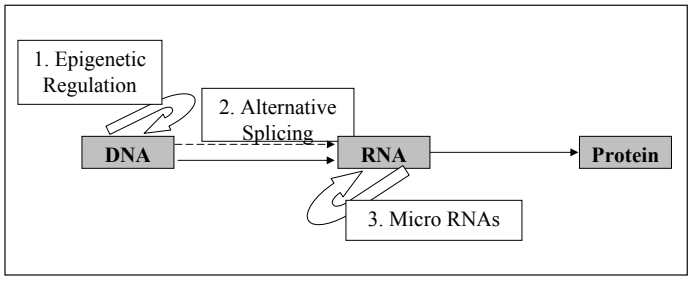


**Figure 1B.**

**Figure 1.** The Central Dogma of Biology. A. (above). As originally put forth, held that one gene at the DNA level is transcribed into one RNA molecule that is translated into one protein. B. (right). A more current view of the Central Dogma accounts for chemical modification of DNA modifying the transcriptional activity of a gene through epigenetic mechanisms (1). Further regulation at the RNA level occurs: (2) by alternative splicing of a transcript diversifying the possible mRNAs that can arise from a single gene or (3) the activity of micro-RNAs upon target transcripts such that a RNA regulates a RNA.

**Non-classical Modes of Inheritance**

Epigenetics is the non-Mendelian inheritance of a trait, or inheritance that is not dependent on the passage of genetic sequence from parents to offspring (Fig. 1B). There are numerous mechanisms of epigenetic gene regulation; one example involves heritable changes to the chemistry of genomic DNA by methylation of cytosine residues. C-methylation has great effect when it occurs within the genomic DNA immediately upstream of a gene known as the promoter. Methylation within promoters decreases the binding affinity of proteins necessary for the synthesis of mRNA resulting in decreased levels of mRNA and the resulting protein (2). Two other examples of epigenetic regulation are post-translational histone modification (3) and the deposition of histone variants in genomic DNA (4). Either of these can change the availability of genomic DNA for binding of the regulatory and/or transcription machinery and result in altered mRNA levels.

A second example of non-Mendelian inheritance is the presence of a regulatory pattern based on the parental origin of the gene in question. This mode of inheritance has been demonstrated in numerous instances. The most general of these is known as imprinting, where gene expression is monoallelic due to transcriptional repression of one copy of the gene in question. This mechanism has been shown to regulate genes involved in development, neural disorders (reviewed in refs. 5-7) and cancer (reviewed in refs. 8-10). A more specific case of imprinting is gene silencing by X-chromosome inactivation in mammalian females (11). This regulation is necessary because females have two X-chromosomes whereas males have one. X-chromosome inactivation turns off expression from one X-chromosome, thereby decreasing X-chromosome derived transcript levels to normal. Transcription from both chromosomes in mammalian females is very detrimental to the organism.

A more recently discovered form of gene regulation is the binding of a small RNA, known as a micro-RNA, to an mRNA in a sequence specific manner (reviewed in (12)and Fig.1B). Micro-RNA binding occurs in a complex consisting of the mRNA, the micro-RNA, and a protein complex known as the RNA-induced silencing complex (RISC). Binding of micro-RNAs to mRNAs leads to gene down-regulation by one of two mechanisms. In the first mechanism, binding of the micro-RNA/RISC complex to mRNA sterically prevents ribosomes from translation of the mRNA (Fig. 2A). The micro-RNA sequence in this regulatory mechanism is not completely complementary to the mRNA sequence. Alternatively, binding of the micro-RNA/RISC complex targets the transcript for destruction by a nuclease in the RISC complex. This mechanism requires that the micro-RNA sequence be perfectly complementary to the mRNA. Micro-RNA based gene regulation has been demonstrated in taxa from yeast and fungi to mammals and is widespread in plants (reviewed in ref. 13). In *A. thaliana,* a mustard plant, more than 1.5 million micro-RNAs have been characterized. The human genome is estimated to encode 400 micro-RNAs that regulate ~5000 genes. This mode of regulation has been shown to be important in the development of almost all human tissues, and recent reports have demonstrated a role for micro-RNAs in cancer (14) and neurological disorders (15). Clearly, micro-RNA gene regulation mechanisms greatly expand the Central Dogma in that an RNA molecule regulates an RNA transcript.

**Staying Ahead of the Curve**

Medical science is rapidly progressing toward the day when "personalized medicine" will be the norm. Genetic analysis of individual patients is at the forefront of this evolution. The Cancer Genome Anatomy Project (http://cgap.nci.nih.gov/; accessed January 2006) is a publicly available data warehouse containing gene expression profiles, chromosomal gene mapping information and tools used in cancer research. Information being gathered there compares expression profiles and sequence information from cancerous, precancerous, and normal tissues in an effort to further characterize target molecules against which to design cancer treatments.

Biomarkers have been defined by the National Institutes of Health as "…an indicator of a disease process, and could replace hard clinical end points as a measure of the effect of new therapies" (16). This document is a request for grant applications to develop biomarkers for "well-defined human diseases of liver, kidney, urological tract, digestive and hematologic systems, and endocrine and metabolic disorders, diabetes and its complications, and obesity" indicative of the wide array of diseases that biomarker discovery is expected to impact. Gene expression research by microarray is a primary component of biomarker discovery because it is high throughput, scanning thousands of candidate genes in a single experiment.

Another use of microarray technology is to identify single-base differences (Single Nucleotide Polymorphisms; SNPs) as well as other genetic polymorphisms (insertions, deletions, or duplications) in a gene sequence. Roche is utilizing this capability in a microarray that screens for a panel of polymorphisms in the cytochrome P450 family of drug-processing proteins (17). Polymorphisms in this family of proteins result in differing drug metabolism rates.

It is a common hypothesis that in the future, genetic profiles will be used as a preventative or predictive measure of a disease state. The Food and Drug Administration has released a guidance document detailing how

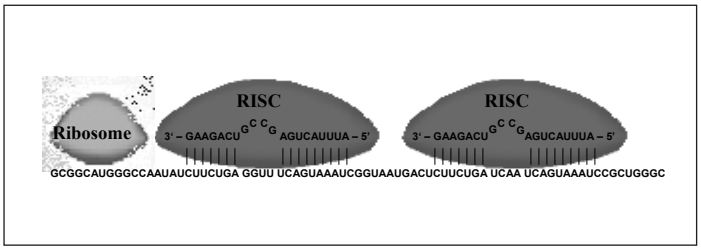


**Figure 2A.**

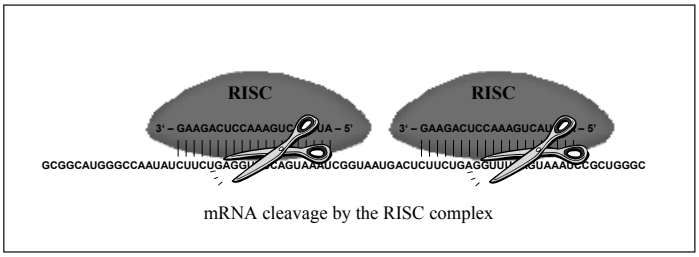


**Figure 2B.**

**Figure 2.** Translational regulation by the RISC complex.
A. Binding of the RISC complex physically blocks the ribosome from translating the mRNA into protein (see text).
B. RISC complex binding results in mRNA degradation (see text).

and when genetic testing or gene expression data may be used in applications for licensing of new drugs (18). This would breach a major regulatory obstacle to drug development for patients exhibiting specific genetic profiles. The marketplace is driving this change in the way drugs are developed by collecting genetic data to find target molecules and then screening huge repositories of small molecules for biological activity against these new targets. Recognition of these market forces is found in the guidance document (18), "...some pharmacogenetic tests - primarily those related to drug metabolism - have well-accepted mechanistic and clinical significance and are currently being integrated into drug development decision making and clinical practice." Health care prognosticators foresee a day when an individual will carry his or her genetic profile on a computer-readable card or chip. Physicians will prescribe therapies specifically tailored to treat disease based on polymorphisms in that individual's genetic code.

**Marker Discovery Projects**

Two projects are underway to survey sequence variability within the human population. The first, taking place outside the United States, is a determination of sequence variability in large national cohort groups. The head of the Human Genome Project, Francis Collins, recently called for a comparable stateside effort to survey human heterogeneity at the sequence level within the United States (19) as a means of determining how environmental factors interplay with genetic sequence in disease. He proposed that the DNA sequencing capacity, originally focused on decoding the human genome, be applied to sequencing a human multi-thousand cohort from all ethnicities to determine the underlying genetics of American phenotypic variability. Additional background information of all types would be collected and used to find corollaries between disease and genetics, based on factors such as diet, lifestyle and exposure to environmental factors.

The second survey, the HapMap project, is an international effort to investigate human heterogeneity. The first phase has recently been published (20). To achieve the goals of the project, common SNPs were sequenced from 269 people who originated from four geographically constrained areas around the world. The goal of the project was "to create a resource that would accelerate the identification of genetic factors that influence medical traits."(20) Phase II of the project will delve into rare SNP sequence variability in the same population (20). The results from these efforts will allow medical professionals to treat disease based on the genetic code of the patient.

Indicative of the applicability of results from basic science to the clinic is the recent announcement that the cancer drug Camptosar® from Pfizer will soon be re-labeled to reflect additional risks associated with a specific mutation in the UGT1A1 gene (21), and the lung cancer drug IRESSA will be marketed heavily in Asian populations reflecting its increased efficacy in that population (22). These two instances highlight the growing trend toward tailoring specific treatments to genetic backgrounds. While it is less likely that gene sequence information will ever have broad applicability in an aviation setting, biomarker discovery has clear and present application to aerospace medicine for a variety of environmental stimuli including fatigue and hypoxia.

## GENE EXPRESSION AT THE FAA

The Federal Aviation Administration is a regulatory agency responsible for aviation safety. In addition to its training and regulatory responsibilities, it is responsible postmortem analysis of accident victims. As medical science enters the post-genome-sequencing era, the agency has the responsibility to investigate new technologies that can lead to better forensic testing, regulation, and certification. Within the Functional Genomics group at the FAA's Civil Aerospace Medical Institute (CAMI) in Oklahoma City, this effort has begun at the level of discovering gene expression changes in response to factors that affect aviation safety. The most efficient method currently available for screening the entire transcriptional output (transcriptome) of a tissue is microarray analysis.

Nucleic acid microarrays quantitatively detect binding of nucleic acids with complementary sequence. Published studies show that microarray analyses are sensitive enough to differentiate not just between normal and disease states, but also between different subtypes of a complex disease with multiple presentations such as leukemia. In addition to Single Nucleotide Polymorphism detection, sequence changes of one base, a common use of microarray technology has been to determine differences in gene expression patterns between disease and healthy states. Therefore, while genetic background cannot be ignored in the search for therapeutic targets, expression information yields tremendous insights into the physiological response to an environmental stimulus. As an expression screening tool, microarrays determine changes in gene expression or the function of the regulatory machinery, not sequence variability. The ability of microarrays to detect subtle expression changes bodes well for our goal to characterize physiologically heterogeneous human factors such as fatigue and hypoxia or exposure to moderate levels of radiation, toxins, or microbes.

### Workflow

Gene expression analysis experiments begin by isolating the RNA from test subjects (Fig. 3). Most studies performed at CAMI use human subjects so the RNA source is whole blood due the simplicity of collection and the demonstrated efficacy of whole blood as a tissue that is responsive to conditions affecting other solid tissues (23-28). A commercial whole blood collection and mRNA stabilization system from PaxGene (PreAnalytiX, Hombrechtikon, CH) is used for sample collection. After mRNA purification, a series of molecular biology steps yields a set of target molecules that maintains the relative gene expression levels of the subject at the time of collection. During synthesis, the target molecules are biotin-labeled for post-hybridization detection with streptavidin/phycoerythrin. Streptavidin tightly binds biotin with very high affinity and specificity. Phycoerythrin fluoresces in response to excitation with light. The system allows the hybridization step to be separated from the detection step such that the bulky fluorescent molecule does not interfere with hybridization between the target and the probe.

Target molecules are hybridized to the arrays in an overnight incubation step. After a washing and SAPE labeling protocol is performed, the arrays are visualized on a fluorescent scanner and the data collected on a computer. The data from these experiments are the raw fluorescence intensities from the phycoerythrin. Genes with higher expression levels in the sample exhibit higher signal intensities on the array. The Affymetrix system differs from other microarray systems in that each sample is individually hybridized to a single array. Most other systems utilize a competitive hybridization wherein two samples, i.e., one each from a normal and a disease subject, are hybridized to an array. Each of the two samples has a different color fluorescent dye conjugated to it. As a result, the data from these arrays is a comparative analysis of the relative fluorescence intensities of each sample against the other. This method has several technical disadvantages related to the different spectral characteristics, stabilities, and incorporation rates of the two dyes.

### Array Design

The arrays are made up of probe sequence covalently attached to a solid substrate, usually glass or silica (Fig. 4). Each probe sequence in a particular location on the substrate is from a single transcript, as annotated in an external sequence database. Arrays with probes representing all human genes encoding known or putative proteins and functional RNAs are available from many commercial sources. Our lab uses arrays from Affymetrix (Santa Clara, CA) due to their ease of use, demonstrated

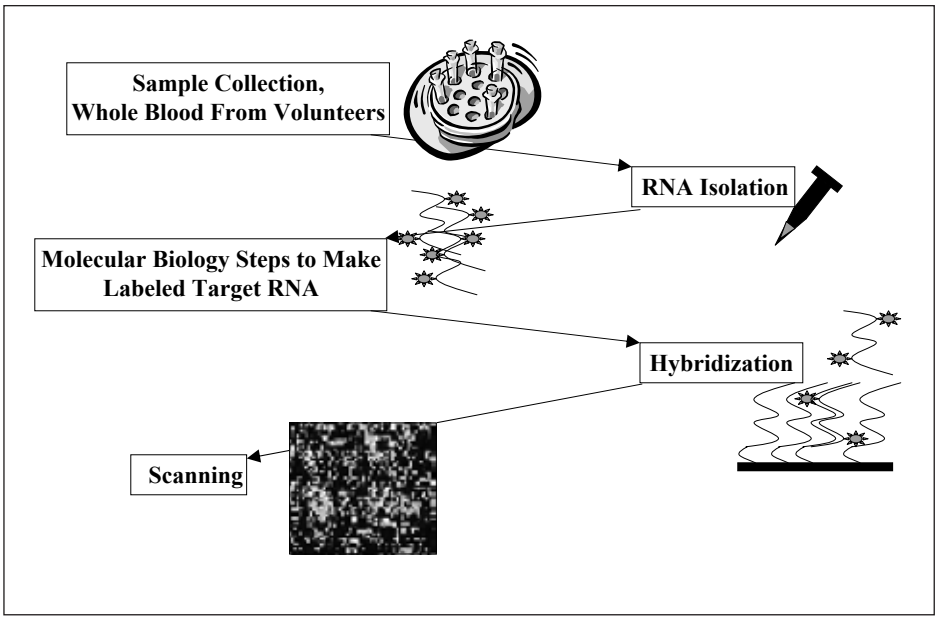


**Figure 3.** Microarray analysis workflow. Samples, usually blood, are collected from volunteer study participants. Total RNA is purified, amplified to make target material, hybridized to chips and the data gathered by scanning the chips.

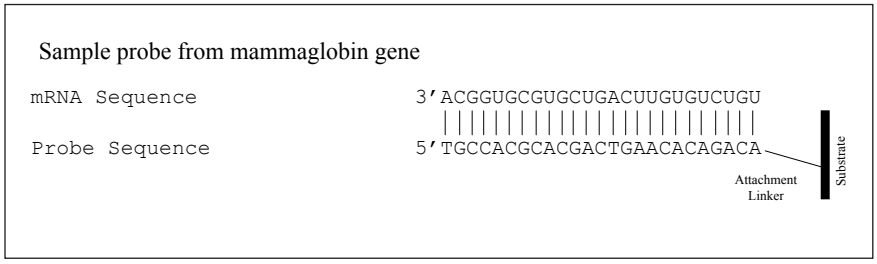


**Figure 4.** Affymetrix GeneChip® design. The Affymetrix GeneChip® design strategy involves multiple short probes to a single transcript instead of a single longer probe to each transcript as with most other manufacturers. The redundancy of the Affymetrix strategy allows the system to detect expression even if hybridization to a single probe is not efficient.

reproducibility of results, and high data quality. Current array architectures are based on either detection of an entire gene or detection of each exon within a gene. Both architectures utilize multiple probes to the sequence of interest, whether it is a gene or just an exon. Both chips can be used to detect an entire gene, as well. The difference is that the exon arrays gather data from four probes to each exon and detect expression at the exon level. The gene arrays interrogate a gene across the entire length of the transcript but lack the redundancy of the exon arrays at the exon level making them unsuitable for exon-level expression analysis. As an example, TMEM8, a gene that was differentially expressed in our study of moderate alcohol use, has 15 exons. The Affymetrix Human Exon 1.0 ST array interrogates these 15 exons with 28 probe sets and comprising 108 probes. The probe set for the same gene on the Human Gene 1.0 ST arrays includes 33 probes, an average coverage of just over two probes per exon.

In addition to differences in probe density, the two array types were designed with different philosophies based on the transcripts that were included in their design. The gene arrays interrogate only well annotated genes to the species of interest. The human Gene1.0 ST arrays have about 28,000 probe sets from known human genes with several sources of validation as to their identity. Conversely, the Exon1.0 ST arrays are much more experimental in nature. Probe sets included in these arrays can be based on as little as sequence homology with a known gene from another organism, even if that gene has never been shown to be active in the species of interest. As a result, there are numerous hypothetical genes included on the exon arrays that are not included on the gene arrays. Clearly, the two chips are designed to answer very different questions necessitating careful experimental design.

**Analysis**

There are numerous sources of signal variation on a microarray. There is signal from the array substrate and from non-specific hybridization of target molecules to probe. Sample to sample variation in signal intensity arises from differences in RNA purity, efficiency of the enzymatic steps to make target, and hybridization efficiency of target to probe. Finally, differences in signal intensity are due to the parameter the experiment is designed to measure, a biological difference in the gene expression level that is attributable to the administered treatment or condition of interest. Further confusing the unambiguous identification of genes that are differentially expressed between treatment groups is the considerable biological variation between individuals.

The raw data consists of total signal intensity values from each individual probe. A pre-processing step to "summarize" the data is performed. The function of this step is to subtract the background intensities, estimated from signal on a blank chip region, and then condense or "summarize" the data from each set of probe into a signal intensity value across the probe set (Fig. 5). The exon array software can return summarization values at both the exon and the gene level.

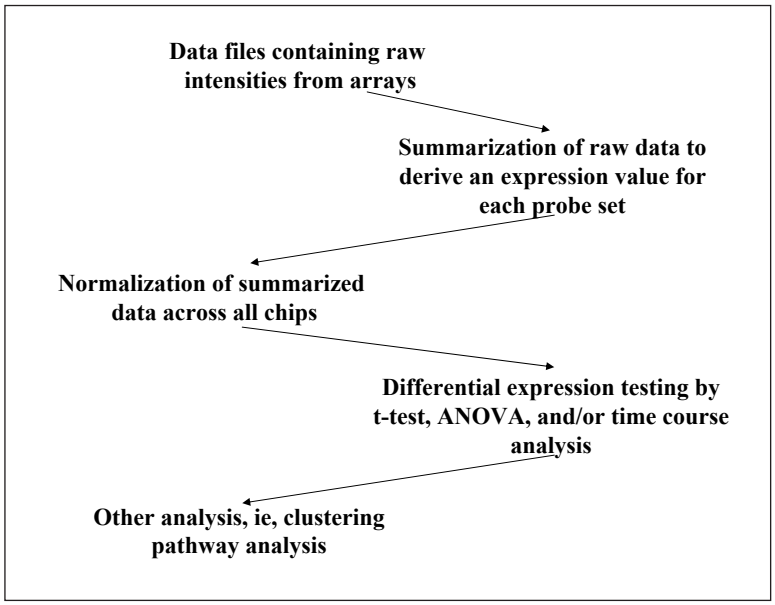


**Figure 5.** The analysis workflow used with Affymetrix GeneChips® (see text).

Background/summarized probe set intensities are then "normalized." Normalization methods computationally adjust signal intensities across the experiment such that between arrays, most signal intensities are statistically equal. The effect of normalization is to adjust the data for biological diversity between subjects or samples. Therefore, after normalization, differences in signal intensity are due to treatment related changes in expression (29). Normalization has been shown to increase the number of genes called differentially expressed over non-normalized datasets.

Due to the normal biological variation between individuals, statistical tools such as t-tests and ANOVA are used to determine which genes are differentially expressed between treatment groups. Because microarrays test the expression of tens of thousands of genes at a time, the statistical analysis is computationally intensive. Specialized software has been developed to perform these analyses on microarray data and deliver an output that reflects biology and is statistically valid despite the necessarily small sample sizes. Furthermore, downstream "data mining" tools such as clustering and pathway analysis have been developed that allow investigators to find genes that change expression levels as a group and can be expected to be similarly regulated and/or be part of the same biological pathway. We use ArrayAnalyzer (Insightful Corporation, Seattle WA) and BioConductor (www.bioconductor.org) for analysis of our data from the Affymetrix system. This software has shown itself to be capable of handling large datasets, limited only by available computational power. Pathway Analysis (Ingenuity Systems, Inc.; Redwood City, CA), a Internet-based knowledge environment, is used to group differentially expressed genes by biological pathway.

Early in the development of microarray technology, microarray results were found to be somewhat irreproducible from lab to lab (29-31). However, recent carefully controlled experiments have shown some improvement (30-32), and inconsistency is thought to be due to the gene regions being interrogated in different platforms. In any event, other molecular biology techniques are used to verify the results. Real-time quantitative polymerase chain reaction (Q-PCR) is the method of choice for initial validation of microarray data and is widely used for initial validation of microarray results. It is simple to use, very reproducible and, as the name implies, can be quantitative with respect to the transcript copy number in the starting sample.

Once a set of genes that change in response to a particular factor has been identified by microarray and validated by Q-PCR, the gene products will be investigated further. Initially, a change in protein level would be confirmed using methods that could include protein arrays, or two-dimensional gels. Further work to investigate pathways affected by a particular stimulus would be performed by under- or over-expression in tissue culture or animal models.

As discussed above, multiple modifications to the Central Dogma have recently been discovered. The analysis technique outlined here is limited to detecting poly-A containing mRNA expression. However, methods are becoming available to assess the effects of other genetic regulatory mechanisms and could be utilized at the FAA with only minor modifications to the workflow detailed above. Currently, two methods for the detection of micro-RNAs are available. A quantitative PCR assay specific for micro-RNAs from Applied Biosystems is newly available and Affymetrix markets whole genome tiling arrays that enable interrogation of the entire genome for novel regions of expression without reliance on an outside database of expression information. Rather, a probe from both DNA strands representing the genomic sequence at predetermined intervals is bound to the array substrate. These arrays require a modification in the amplification of target material; random primers are used for amplification of total RNA in order to amplify all RNAs, irrespective of the presence of a poly-A tail. Other methods that take advantage of the tiling arrays for detection but utilize alternative methods for target isolation are available for discovery of regions of epigenetic control.

## CURRENT GENE EXPRESSION STUDIES

The aviation accident rate has been steadily declining over the last several decades. Senior FAA leadership recognizes that further decreases in the accident rate will be incrementally more difficult than what has been achieved to date. FAA employees have been challenged to become ever more creative in their approach to improve aviation safety. A determination of gene expression profiles in response to human factors that affect aviation safety could lead to better fact-based regulations. The area of forensics could benefit greatly from determining the gene expression or metabolic profile of related factors with the goal of being able to assign causation to accident investigations. In addition, there are issues that affect the working environment of aircraft personnel that particularly lend themselves to investigation by microarray or other genetic methodologies. Experimental design typically includes human performance data in order to link gene expression to performance levels.

### Moderate Alcohol Use

We are studying the effects of moderate blood alcohol levels on gene expression. A current study in collaboration

with the University of Utah and The Center for Human Toxicology is looking at expression changes in response to blood alcohol levels up to 0.08%, the legal limit for automobile operation in most states. Data from six men in their early 20's have been collected, and preliminary data analysis has found about 400 genes that change expression level at some point between baseline, 0.08%, and return to 0.00% blood alcohol. Further data analysis and verification are underway, and additional experiments are being planned to expand this work beyond the current data set.

**Fatigue**

The National Transportation Safety Board is very interested in fatigue as a potential causative factor in accidents from all forms of transportation accidents, not just aviation. In addition, there is support at both the state and federal government levels to investigate fatigue as a factor in automobile safety. In collaboration with the U.S. Air Force, Brooks City-Base in San Antonio, Texas, we are investigating the effects of fatigue on gene expression. This study compares gene expression levels at up to 35 hours of sleeplessness with a well-rested baseline.

**Radiation**

A condition that is somewhat unique to the aviation industry is cosmic radiation. Passengers have not been shown to be at risk due the short duration of increased exposure to radiation during a single flight. However, there is evidence that flight crews have an increased incidence of chromosomal aberrations, melanomas and breast cancers (33-36), presumably from increased exposure to solar radiation. Other researchers have demonstrated changes in gene expression levels due to ionizing radiation (37-41). A recent study discovered different sets of markers to three levels of ionizing radiation exposure in mice (42).

**Hypoxia/Altitude Exposure**

Decreased blood oxygen, or hypoxia, is a human factor that also is unique to the aviation industry. As such, we are positioned to investigate changes in gene expression from hypoxic conditions that may arise due to increasing altitude. The necessary equipment to mimic increasing altitude conditions is available at the CAMI. One current study is investigating the effects of exposure to 12,400 feet in a male population. A second study at airline cabin altitudes of 6,000 and 7,000 feet is in the planning stages.

## GENOTYPING STUDIES

Within the Functional Genomics group, a protocol has been developed (43)to genotype accident samples using the Agilent BioAnalyzer to separate PCR product from the Federal Bureau of Investigation's approved CODIS marker set (44). Others have used this instrument to genotype papillomavirus (45), markers on the human Y chromosome (46), and markers of mitochondrial disease (47). Our capabilities are limited to comparative analyses; nonetheless, this ability is useful to the Forensic Toxicology Research Team on several fronts. First, it is useful for confirming sample identity in cases where samples are co-mingled at the accident site. Second, genotyping has allowed the confirmation of chemical toxicology results that found distinct differences between two samples that were documented as having been from the same accident victim. Third, the ability to perform this level of genotyping "in house" is a great cost saver to the federal government; this assay performed on the BioAnalyzer is done at one-tenth the cost of sending it out to a forensic genotyping laboratory.

## FUTURE STUDIES

All of the work in progress in the Functional Genomics lab at CAMI is among the first to investigate the affects of these particular aerospace environmental factors on gene expression. Before any global conclusions can be reached, each of these studies will need to be extended and validated. As an example, the fatigue study is being performed with samples from Air Force personnel who have all passed flight physicals. Will the data gathered from this study translate broadly across genders or to an older and/or less fit population? Candidate genes discovered by microarray analysis on limited patient populations will be further explored by Q-PCR on groups of subjects from a population base that more generally reflects the overall population.

The power of the available software to find genes that are regulated coordinately will allow us to explore entire biological pathways. In this way, genes that may not appear by microarray analysis but play a role in the factor being investigated will become visible. Tissue culture experiments are expected to play a role in this work because they will allow for cells from many different tissues to be investigated, thereby expanding the work into tissues other than blood. Multiple avenues of investigation become available in tissue culture. Overexpression of genes of interest in tissue culture can then be extended into animal models to confirm the findings *in vivo*.

We expect that in response to the factor of interest, there will be some genes that are up-regulated and others will be down-regulated. A molecular signature for a particular factor will therefore consist of multiple genes and their interrelationships. In addition, the factors we are interested in exploring are biologically not on or off but are expected function on a continuum. Therefore, multiple signatures are needed to determine where on the continuum an unknown sample lies (Fig. 6).

RNA is an especially labile molecule. As such, assays for the more stabile gene products (proteins) are likely to be the end goal of this body of work. A potential assay for forensic testing is a custom protein array that would assay protein levels of a subset of proteins selected for their specificity across a range of conditions of interest. A sample would be brought into the lab and hybridized to the array. The signature seen in the unknown sample would be compared with normals and the known signature continuums for all known human factors. The unknown sample would be categorized as being either normal or impaired and the cause of the impairment determined in this way.

In the post 9-11 world, terrorism has moved to the forefront of our national consciousness. Within functional genomics, numerous avenues of investigation present themselves in the war on terror. While opportunities to develop pathogen identification methods exist, the real power of gene expression technology can be brought to bear on two fronts. First, by determining the gene expression changes within the pathogen, new angles to attack the organism and treat the infection can be determined. Second, host gene expression changes that occur post-infection can be used to discover new host target pathways for treatment of the symptoms of the infection. Knowledge of gene expression changes along the infection time-line would allow clinicians to triage patients and prioritize treatment to those who are at greatest risk. Finally, most of these same questions could be answered for exposure to chemical agents.

## CONCLUSIONS

The FAA is responsible for public safety through regulation and oversight of the aviation industry. There has been a dramatic decrease in accident rates to the point that American skies are the safest in the world. However, accidents still occur and it is the goal of the agency to further decrease the accident rate. To that end, CAMI has been tasked to investigate the human equation from all angles as it pertains to aviation. The Functional Genomics group at CAMI was formed in part to bring aviation medicine into the molecular age and discover the unique molecular profiles that result from an aerospace safety-related stimulus. Factors we are investigating or will investigate have been shown to impact aviation safety. Factors such as fatigue and hypoxia are of great interest in all branches of aviation; alcohol can be a factor in general aviation. In addition, cosmic radiation poses a long-term risk to the health and well being of pilots and aircrew, especially as space flight becomes reality. An increased understanding of these factors at the molecular level will increase safety for the flying public and aircrews through better regulation, development of targeted therapeutic interventions and increased understanding of the risks posed by flight in general.

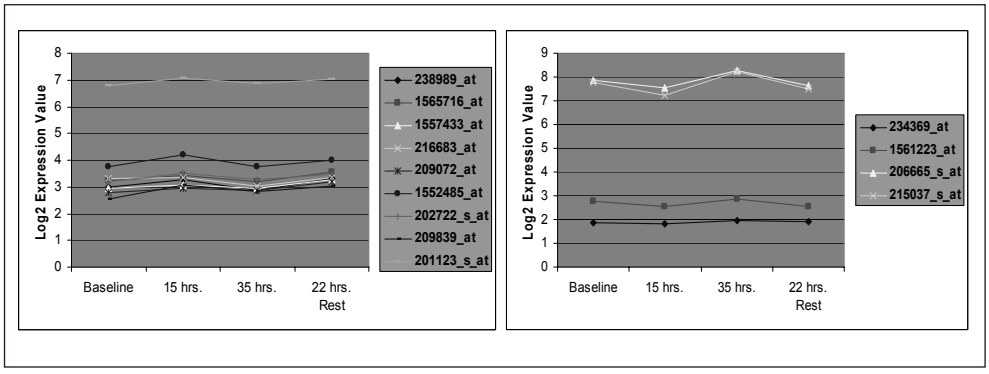


**Figure 6.** Resolution of a set of differentially expressed genes into two expression patterns, or clusters. Clustering of differentially expressed genes demonstrates two expression patterns which, when taken together resolve into a cohesive whole that can be used to differentiate between the timepoints in this study of sleeplessness. Addition of time points and correlation of expression data to performance data will fine-tune the ability of the assay to differentiate among levels of impairment.

## REFERENCES


(1) Perry GH, Dominy NJ, Claw KG, et al. Diet and the evolution of human amylase gene copy number variation. Nat Genet. 2007;39(10):1256-60.

(2) Wolffe A. DNA methylation and chromatin. *Chromatin, structure and function*, 2 ed. San Diego, CA: Academic Press 1995:94-6.

(3) Wolffe A. Post-translational modification of core histones. *Chromatin, structure and function*, 2 ed. San Diego CA: Academic Press 1995:72-7.

(4) Wolffe A. Histone variants. *Chromatin, structure and function*. 2 ed. San Diego, CA: Academic Press 1995:68-72.

(5) Soejima H, Wagstaff J. Imprinting centers, chromatin structure, and disease. J Cell Biochem. 2005 May 15;95(2):226-33.

(6) Keverne EB. Genomic imprinting in the brain. Curr Opin Neurobiol. 1997 Aug 7;(4):463-8.

(7) Cassidy SB, Schwartz S. Prader-Willi and Angelman syndromes. Disorders of genomic imprinting. Medicine (Baltimore). 1998 Mar;77 (2):140-51.

(8) Gerald WL. The molecular genetics of Wilms tumor: A paradigm of heterogeneity in tumor development. Cancer Invest. 1994;12(3):350-9.

(9) Chung SW, Daniel R, Wong BY, Wong P, M. The ABL genes in normal and abnormal cell development. Crit Rev Oncog. 1996;7(1-2):33-48.

(10) He H, Olesnanik K, Nagy R, Liyanarachchi S, Prasad ML, et al. Allelic variation in gene expression in thyroid tissue. Thyroid. 2005 Jul;15(7):660-7.

(11) Wolffe A. Heterochromatin, position effect, locus control regions and insulators. *Chromatin, structure and function,* 2 ed. San Diego CA: Academic Press 1995:84-94.

(12) Zamore PD, Haley B. Ribo-gnome: The big world of small RNAs. Science. 2005 Sep 2;309(5740):1519-24.

(13) Vaughn MW, Martienssen R. It's a small RNA world, after all. Science. 2005 Sep 2;309(5740):1525-6.

(14) He H, Jazdzewski K, Li W, et al. The role of microRNA genes in papillary thyroid carcinoma. PNAS. 2005 Dec 19;19075-80.

(15) Kosik KS, Krichevsky AM. The elegance of the microRNAs: A neuronal perspective. Neuron. 2005 Sep 15;47(6):779-82.

(16) Development of Disease Biomarkers. 2005 March 14, 2006 (cited 2007 October 15); Program announcement. Available from: grants.nih.gov/grants/guide/pa-files/PA-05-098.html.

(17) Roche diagnostics. Background information: The technology behind Amplichip microarrays. *Technical Bulletin*. Basel, Switzerland.

(18) Guidance for industry: Pharmacogenomic data submissions. [URL] 2005 [cited 2005 March 22, 2005]; Available from: www.fda.gov/cber/gdlns/pharmdtasub.htm.

(19) Collins FS. The case for a US prospective cohort study of genes and environment. Nature. 2004 27 May; 429: 475-7.

(20) The International HapMap Consortium. A haplotype map of the human genome. Nature. 2005 27 Oct;437:1299-320.

(21) NDA 20-571 Camptosar®(irinotecan HCl); Final label. U.S. Food and Drug Administration 2005:17.

(22) New hope for cancer patients: Iressa (gefitinib) launched in China for treating advanced non-small cell lung cancer. 2005 [cited 25 Feb 2005]; Available from: http://en.astrazeneca.com.cn/Article/511765.aspx.

(23) Pahl A, Brune K. Gene expression changes in blood after phlebotomy: Implications for gene expression profiling. Blood. 2002;100(3):1094.

(24) Hakonarson H, Bjornsdottir US, Halapi E, et al. Profiling of genes expressed in peripheral blood mononuclear cells predicts glucocorticoid sensitivity in asthma patients. PNAS. 2005 11 Oct;102(41):14789-94.

(25) Connolly PH, Caiozzo VJ, Zaldivar F, et al. Effects of exercise on gene expression in human peripheral blood mononuclear cells. J Appl Physiol. 2004 1 Oct;97(4):1461-9.

(26) Lee MS, Hanspers K, Barker CS, et al. Gene expression profiles during human CD4+ T cell differentiation. Int Immunol. 2004 1Aug;16(8):1109-24.

(27) Nicholson AC, Unger ER, Mangalathu R,et al. Exploration of neuroendocrine and immune gene expression in peripheral blood mononuclear cells. Brain Res Mol Brain Res. 2004;129:193-7.

(28) Qing X, Putterman C. Gene expression profiling in the study of the pathogenesis of systemic lupus erythematosus. Autoimmun Rev. 2004 3 Nov; (7-8):505-9.

(29) Draghici S. Array normalization. *Data analysis tools for DNA microarrays*. New York, NY: Chapman & Hall/CRC 2003:313-8.

(30) Larkin JE, Frank BC, Gavras H, et al. Independence and reproducibility across microarray platforms. Nat Methods. 2005 2 May; (5):337-44.

(31) Members of the Toxicogenomics Rersearch Consortium: Weis BK. Standardizing global gene expression analysis between laboratories and across platforms. Nat Methods. 2005 2 May; (5):351-56.

(32) Irizarry RA, Warren D, Spencer F, et al. Multiple-laboratory comparison of microarray platforms. Nat Methods. 2005 2 May; (5):345-50.

(33) Hagmar L, Stromberg U, Bonassi S, et al. Impacts of types of lymphocyte chromosomal aberrations on human cancer risk: Results from Nordic and Italian cohorts. Cancer Res. 2004;64:2258-63.

(34) Romano E, Ferrucci L, Nicolai F, et al. Increase of chromosomal aberrations induced by ionising radiation in peripheral blood lymphocytes of civil aviation pilots and crew members. Mutat Res. 1997;377:89-93.

(35) Hagmar L, Brugger A, Hansteen I-L, et al. Cancer risk in humans predicted by increased levels of chromosomal aberrations in lymphocytes: Nordic study group on the health risk of chromosome damage. Cancer Res. 1994 1 Jun;54(11):2919-22.

(36) Hagmar L, Bonassi S, Stromberg U, et al. Chromosomal aberrations in lymphocytes predict human cancer: A report from the European Study Group on Cytogenetic Biomarkers and Health (ESCH). Cancer Res. 1998 15 Sep;58(18):4117-21.

(37) Grace MB, McLeland CB, Blakely WF. Real-time quantitative RT-PCR assay of GADD45 gene expression changes as a biomarker for radiation biodosimetry. Int J Radiat Biol. 2002;78(11):1011-21.

(38) Amundson SA, Fornace Jr. A.J. Gene expression profiles for monitoring radiation exposure. Radiat Prot Dosimetry. 2001;97(1):11-6.

(39) Amundson SA, Do KT, Fornace Jr. AJ. Induction of stress genes by low doses of gamma rays. Radiat Res. 1999;152:225-31.

(40) Amundson SA, Do KT, Shahab S, et al. Identification of potential mRNA biomarkers in peripheral blood lymphocytes for human exposure to ionizing radiation. Radiat Res. 2000;154:342-6.

(41) Amundson SA, Bittner M, Meltzer P, et al. Induction of gene expression as a monitor of exposure to ionizing radiation. Radiat Res. 2001;156:657-61.

(42) Dressman HK, Muramoto GG, Chao NJ, et al. Gene expression signatures that predict radiation exposure in mice and humans. PLoS Medicine. 2007 1 Apr;4(4):e106.

(43) Kupfer DM, Huggins M, Cassidy B, et al. A rapid and inexpensive PCR-based STR genotyping method for identifying forensic specimens. *FAA Office of Aerospace Medicine Technical Reports*. Washington, DC. 2006; Report no. DOT/FAA/AM-06/14.

(44) Budowle B, Moretti TR, Niezgoda SJ, Brown BL. CODIS and PCR-based short tandem repeat loci: Law enforcement tools. Second European Symposium on Human Identification. Promega Corp; 1998;73-88.

(45) Liu CH, Ma WL, Shi R, et al. Application of Agilent 2100 Bioanalyzer in detection of human papilloma virus. Di Yi Jun Yi Da Xue Xue Bao. 2003 23 May; (3):213-5.

(46) Jabasini M, Zhang L, Dang F, et al. Analysis of DNA polymorphisms on the human Y-chromosome by microchip electrophoresis. Electrophoresis. 2002;23(10):1537-42.

(47) Lu CY, Tso DJ, Yang T, Jong YJ, Wei YH. Detection of DNA mutations associated with mitochondrial diseases by Agilent 2100 bioanalyzer. Clin Chim Acta. 2002;318(1-2):97-105.