Microarray analysis of human hepatic aging

A Thesis

Submitted to the Faculty

of

Drexel University

by

David W. Boorman

in partial fulfillment of the

requirements for the degree

of

Master of Science in Biomedical Science

August 2008

Acknowledgements

Special thanks to Dr. Kriete, my advisor, and Dr. McEachron, Dr. Sokhansanj, Dr. Bentz and Dr. Liao of Drexel University for serving on the master's review committee for this project.

Table of Contents

List of Tables

List of Figures

Abstract

Microarray analysis of human hepatic aging
David W. Boorman
Andres Kriete, PhD

Two sets of liver tissue samples of human subjects were investigated by gene expression microarray analysis. One set had normal and disease samples with their telomere lengths; the other set consisted of groups of normal young and old donors, mostly post mortem. The goal was to identify genes related to aging and inflammation and to investigate the hypothesis that telomere length is a biomarker for these processes.

The first set of ten surgical liver biopsies were run on the single channel microarray, Codelink Human Bioarray. Because of the liver diseases in the subjects (3 with Hepatitis C, 3 with Primary Sclerosing Cholangitis, 1 with a liver cyst, and 3 controls), the standard preliminary procedure of simple variance filtering was deemed insufficient. Instead, the samples were adjusted for disease using Analysis of Covariance, to identify genes of potential interest. This was conducted in three pairs based on age, telomere length and disease, because of the small sample size. With this procedure, about 1,600 genes had an alpha significance of less than five percent (14.7% total for all comparisons). Hierarchical clustering of these genes grouped the Hepatitis C patients and the young patients, as a partial validation of this model.

The second dataset was obtained from the Gene Expression Omnibus (GEO), GSE9588. The samples were run on the Rosetta/Merck Human 44k 1.1 dual channel microarray platform, against a control of 191 pooled human liver samples. The original study analyzed the surgical and post mortem liver samples with respect to disease and single nucleotide polymorphisms. Sixty-seven samples were selected from this dataset, ages 20-29 or ages 70-81. Their telomere lengths are unknown. Simple variance filtering resulted in 1,965 differentially expressed genes.

To reduce false positives further, various microarray techniques were employed, including hierarchical clustering, principal component analysis, k-means clustering and Pavlitis Template Modeling. The last

technique matches expression to the relative telomere length or age. Resulting genes were then run through the gene ontology programs DAVID and PANTHER, which find gene networks.

Networks that were differentially up regulated based on age include ubiquitin, a protein intracellular transport marker often linked to protein targeted for degradation, as well as cytochrome p450 (CYP), a mono-oxygenase in the endoplasmic reticulum of hepatocytes involved in the breakdown of steroids, fatty acids and xenobiotics. The cytochrome p450 network was up regulated in the older group, but not in elderly who also had steatosis (fatty liver) or who had drug liver risk from the use of a medication. Networks down regulated or mixed based on age included the broad categories of "receptors" and "immunity and defense." IL8, which is involved in inflammation, was one of these hits. Networks that were down regulated with shorter telomere length included genes related to chromatin restructuring and transcription regulation, including the selective down regulation of several Histone H4 isoforms.

Hypotheses generated from this study include the identification of six poorly characterized receptors which may be involved in liver regeneration. Additionally, there may be a mechanistic problem between CYP transcription and activated protein, because of the high CYP mRNA levels seen here in elderly, compared to the unchanged or decreased protein levels found in other studies. Additionally, elderly with steatosis or drug liver risk may have a higher hepatocyte turn over and higher Hepatocyte Growth Factor (HGF) expression, since HGF suppresses CYP.

INTRODUCTION

Overall Goals

The purpose of this study is to identify genes and gene networks in the human liver which are differentially regulated by biological age. This will help with understanding the aging process, and understanding differences in physiology which might be relevant in the treatment of diseases.

Review of Aging Theories

Theories of why people age can generally be categorized into three major groups: The Good of the Species, Evolutionary Aging Theory and Rate of Living [1, 2].

The Good of the Species theory basically states that if aging did not occur, the necessary turnover of generations would not occur, and the species would not improve, evolve and adapt. The theory's problems include that it confuses death with aging; we could be young and beautiful, and then die. Additionally, the urge of an individual to reproduce would win out evolutionarily to an ideal zero population growth.

The geneticist Haldane theorized that Huntington's Disease, which is caused by a dominant mutation in the huntingtin protein, is less affected by natural selection because Huntington's Disease does not manifest itself until midlife. Medawar expanded this theory to aging, creating the Evolutionary Aging Theory. This theory states that any gene that affects the body only late in life is virtually immune to natural selection.

The majority of gene mutations are harmful, and some genes have multiple effects. For example, testosterone has evolutionary beneficial effects, such as encouraging muscle development, but suppresses the immune system, accelerates the deterioration of arterial walls and increases the risk of prostate cancer. Evolution would tend to encourage higher expression of testosterone, despite these harmful effects; this is known as antagonistic pleiotrophy.

More generally, animals that are more efficient at repelling predators tend to live longer because they do not need to reproduce as quickly to survive. This explains the general trend of longer life among larger animals, but also explains exceptions to this rule, such as birds, bats, the quahog clam, which has the longest life recorded at about two hundred years, and others. Flying squirrels live about 17 years, versus 7-8 years for chipmunks and 2-3 years for rats. Island animals which have fewer predators also tend to live longer than their mainland counterparts.

The proponents of the Rate of Living Theory include Max Rubner and Raymond Pearl. Studies show that in general, larger animals have slower metabolic rates and tend to live longer [3]. Others [2] showed that longevity in rats can be increased by decreasing diet consumption. While there is a general trend with this, there are important exceptions to this rule in a number of small animals with long lives which are not explained by this theory. Birds and bats both have very long lives relative to their relative metabolism. In bats, this was suspected to be caused by their hibernation. However, tropical bats that do not hibernate are just as long lived. Additionally, the total number of heart beats of a water flea are about the same whether in cold water, which slows their heart rate, or warm water. Marsupials have metabolic rates 70-80% of that of other animals the same size, but marsupials are shorter lived. Sacher's animal experiments predict an expenditure of 200-300 kcal/g per lifetime, but humans live about four times as long as predicted, with an expenditure of 800 kCal/g.

The rate of living theory is strongly associated with the generation of reactive oxygen species (ROS) through oxidative metabolism in mitochondria [4]. With the development of molecular biology, this theory has received particular attention in the search for molecular mechanisms behind aging process. Mitochondria are highly dynamic organelles, attracted to cytosolic sites of high ATP demand and undergo constant fission and fusion. However, there is substantial complexity in the process from free radical generation to mitochondrial DNA damage and loss of mitochondrial turnover, normally modulated by antioxidant and repair pathways [5], and further regulated by the mechanisms of uncoupling proteins (UCPs), which belong to a class of mitochondrial inner-membrane proteins [6]. Since ROS is highly

diffusible, damage to mitochondrial proteins and DNA closest to the origin of ROS production are most likely, but oxidation of proteins can occur throughout the cell. Interestingly, it has been suggested that ROS preferentially oxidizes resident proteins in the ER [7], making ROS the primary cause not only for mitochondrial, but also for ER dysfunction.

Telomere Shortening

With the discovery of cellular senescence in-vitro [8], another mechanism contributing to the aging process was discovered, the shortening of telomeres. Telomeres are repeated DNA sequence, GGGTTA in humans, that are about 10,000 base pair long, and found at the ends of chromosomes. Each time the cell divides, the telomere gets 50-100 bases shorter, because the DNA polymerase falls off the end before it completes replication. In the absence of telomerase, which is capable of relengthening the telomeres, the cell will enter senescence when the telomeres become short enough, and the cell will no longer divide. This corresponds with changes in gene expression [9] This serves in part as a control mechanism for carcinogenesis.

Aging and Disease

Disruption of mitochondrial function, generation of ROS and cellular oxidative damage has been associated with chronic inflammation implicated in a host of degenerative disease states including osteoarthritis, atherosclerosis, type-2 diabetes and even cancer [10-12]. Age-associated chronic inflammatory states are distinct from inflammation triggered by infection. It is presently unclear to what extent chronic inflammatory states in older individuals represent autoimmune processes caused by deregulation of the immune system [13, 14]. Alternatively, these states may arise as a consequence of an increased oxidative stress response in old cells triggered by molecular damage incurred over a lifetime. In support of cell-autonomous causes for age-associated inflammation, expression of inflammatory markers such as cytokines, has been observed in cells subjected to replicative senescence in vitro caused by serial passaging [15-18].

Liver and Aging

Although the liver is more regenerative than other organs, age does not spare the liver, in particular because its one of the most metabolically active organs. Both hepatic blood flow and total liver volume decrease with age [19, 20], as does its metabolic activity. As within other cell types, reactive oxidative species damage DNA. Mitochondrial DNA is especially prone, because it has fewer means to correct errors from mutations [21]. Response to disease is also changed; for example, people infected with Hepatitis C virus after 40 years of age will progress more quickly to fibrosis [22].

Steroid and Drug Metabolism in the Liver

A major function of the liver is to metabolize both steroids and xenobiotic molecules. Typically, this is categorized into two phases. Phase I involves the oxidation of the molecule, while Phase II involves conjugation to make the molecule water soluble, so it can be excreted by the kidney. Cytochrome p450 (CYP) is an example of a Phase I protein family, while glutathione s-transferase (GST) is an example of a Phase II protein family. Many drugs are poorly eliminated in the elderly. However, this is not necessarily because of reduced CYP activity. Other physiological changes such as reduced renal capacity, reduced hepatic blood flow and reduced liver volume most likely play a large role [23].

Cytochrome p450

One protein family which became particularly important for this study is cytochrome p450, located in the endoplasmic reticulum. Cytochrome p450 (CYP) are a family of heme-thiolate monoxygenases (adds a single oxygen to a molecule). It is part of an NADPH-dependent electron transport pathway. It oxidizes a variety of structurally unrelated compounds: steroids, fatty acids and xenobiotics. The first three or four families oxidize most xenobiotics, along with other endogenous substrates. Table 1 below lists some families and drugs they metabolize [23, 24].

Table 1: Principal Characteristics of Cytochrome p450 Enzyme Families

Family	Sub-family	Endogenous Biosynthesis	Typical Drug Substrate	Role in Bioactivation (Drugs) or Function (Endogenous)	Inducibility by Chemicals	Isoforms Up Regulated in This Study.
CYP1	A B		Theophylline Theophylline	Very extensive Very extensive	Very high High	CYP1A1 --
CYP2	A B C D E		Propranolol Coumadin Tolbutamide Debrisquine Chlorzoxazone	Limited Limited Poor Poor Extensive	Modest High Modest Not inducible High	2A6, 2A7, 2A13 CYP2B6 CYP2C8
CYP3	A		Erythromycin	Limited	High	CYP3A4, 3A7
CYP5-8		Thromboxane prostacyclin		Biosynthesis		
CYP7	A	Cholesterol		Degradation		CYP7A1
CYP26		Retinoic Acid		Transcription reg.		CYP26A1
CYP51		Cholesterol, Bile acids		Biosynthesis		CYP51A1

Regulation of Cytochrome p450 (CYP) Gene Expression and Transcription

This study involves the use of microarrays, which simultaneously measures the mRNA levels (gene expression) of all 30,000 genes in a sample at once, including 2000 genes related to metabolism, and 70 cytochrome p450 genes (including ten clones or ESTs). This allows us to identify gene networks which are differentially expressed. Regulation of CYP transcription can generally be affected by a variety of factors other than age: genetic background, nutritional status, presence of disease, and previous exposure to other xenobiotics. It is regulated by hormones, such as growth hormones and thyroid hormones. It is induced by alcohol, eating cruciferous vegetables or charcoal-broiled beef. Some families are down regulated by grapefruit juice, calcium blocking drugs or other drugs. Diseases such as hepatocellular carcinoma and insulin-dependent diabetes mellitus (which cause hormonal perturbations) can also change regulation of cytochrome p450 [23].

MATERIALS AND METHODS

Materials and Methods Introduction

Datasets

Two datasets were used in this analysis. The first dataset consisted of ten liver samples obtained and processed by us. The second dataset was obtained from the National Institutes of Health's Gene Expression Omnibus (GEO) database, identification number GSE9588. Both datasets are described in more detail below.

Liver Diseases in These Datasets

Primary Sclerosing Cholangitis (PSC)

PSC is a chronic, slow progressing disease which causes scarring and inflammation of the liver ducts. Scarring leads to bile accumulation, which leads to cirrhosis. Blockage leads to occasional bile infections. Liver transplants are often needed at late stages. The cause is unknown, although about 70% of those affected are men. There is no cure or specific treatment [25].

Hepatitis C

Hepatitis C is a virus transmitted through contact with blood, usually by sharing injecting needles. About 15-40% of those infected are able to clear the infection in the early stages. The remaining develop chronic infections, which can be symptomless for long periods. It can lead to cirrhosis, liver failure and liver cancer. The virus can be treated with a 50% success rate by the antiviral drugs, peginterferon and interferon [25]. However, antiviral treatment is generally not recommended for the elderly, due to a lower rate of response and a higher rate of adverse effects [26].

Liver Cysts

Liver cysts are benign encapsulated growths [25].

Steatosis (GEO GSE9588 Dataset)

Steatosis, or fatty liver, is the buildup of fat within hepatocytes, defined when the fat comprises more than ten percent of the total liver weight. Often there are no symptoms, but it can lead to inflammation of the liver. It can be caused by obesity, diabetes and high triglyceride levels, as well as by alcohol abuse, rapid weight loss and malnutrition. Treatment includes diet change, exercise, control of diabetes, limitation of triglyceride intake, and abstaining from alcohol consumption. [25] A population survey in Italy found steatosis through ultrasound in 16.4% of the general population, 46.4% of heavy drinkers, 75.8% of obese subjects, and 94.5% of heavy drinkers who are obese [27].

MATERIALS AND METHODS

Ten Liver Samples - Description and Data Processing

Tissue Preparation and Microarray Platform

From this study, RNA was isolated from liver tissue obtained from ten live patients who were already undergoing the surgical liver biopsy for medical reasons [28]. Qiagen RNeasy mini kit was used according to the manufacturer's instructions. Gene expression analysis was performed using the Codelink Human Bioarray containing more than fifty thousand single-stranded 30-mer oligonucleotide probes and single dye labeling (Applied Microarrays, Tempe, AZ). Chips were run in duplicate. Details of this platform are available on the vendor's homepage website. Characteristics of the Codelink platform have been previously evaluated by us [29] and as part of the microarray quality control (MAQC) assessment [30]. Sample preparation and hybridization followed procedures previously described [29]. The data were normalized according to the manufacturer's instructions. Average telomere restriction fragment (TRF) length of the hepatocytes was determined by restriction enzyme digestion, gel electrophoresis, and radioactive label hybridization, described elsewhere [28].

Population Demographics

The ten liver tissue samples came from patients aged 21-61, and with telomere lengths 6.2 - 9.0. Since the liver biopsies were performed on live patients undergoing the biopsy for medical reasons, most also had a liver disease, which complicates the analysis. There were three patients with Hepatitis C virus, three with Primary Sclerosing Cholangitis (scarring of the liver ducts), one with a liver cyst, and three with no known liver diseases. To give an overview of the population demographics, we plotted telomere length versus age with their diseases, as shown in Figure 1 below. Ellipses show the consistent grouping of the samples by hierarchical clustering using a number of different distance metrics, discussed further in the Results section.

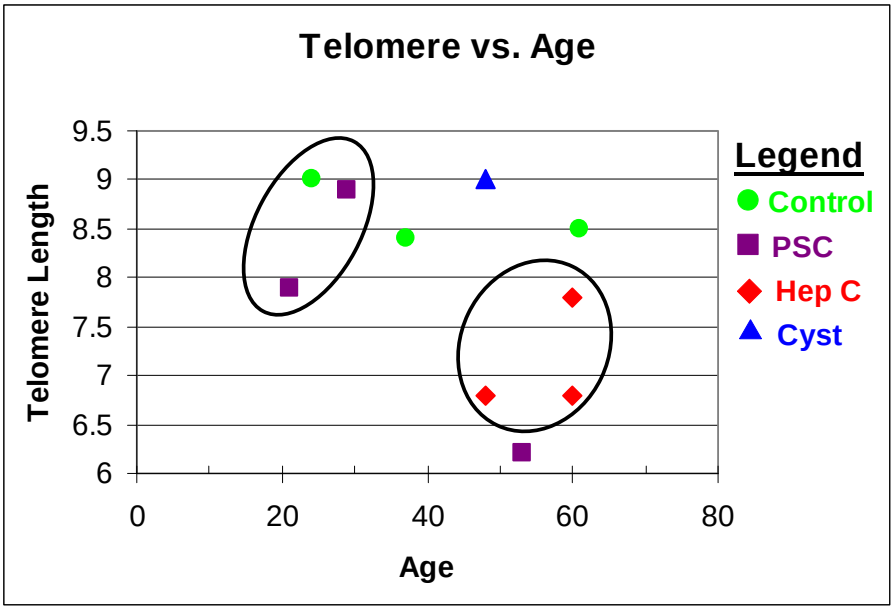

Figure 1: Population demographics of the ten liver samples by average telomere length, patient's age, and by the patient's disease. Hierarchical clustering of the samples consistently grouped the three patients with Hepatitis C and the three patients aged 21-29.

ANCOVA Rationale

Typically with microarray analysis, the standard method in one of the first steps is a broad variance filter of the genes. The vast majority of the genes represented by the 54,000 probes are either not expressed, or are not differentially expressed (aka "housekeeping genes"). This initial filter screens out most of the genes where expression is mostly absent or "flat" across the samples, greatly reducing the size of the dataset, so it can be examined more closely with more sensitive methods. It achieves this by a straightforward check of the variance across the difference samples. Those probes with too little variance are removed.

Because of the small relative sample size and the complexity of the population, this method was deemed inadequate. The presence of three variables (age, disease and telomere length) makes discriminating between them difficult. Equally important, it is likely that a real difference based on age would be masked by noise created from the other variables.

This problem was addressed by adjusting for the diseases using Analysis of Covariance (ANCOVA), instead of simple variance filtering as a preliminary screen of the genes. ANCOVA has been applied

previously to microarray analysis by other researchers, but in the context of a Bayesian framework. [31] ANCOVA is a type of General Linear Model (GLM) which combines multiple linear regression with Analysis of Variance (ANOVA), in an additive fashion. Figure 2 below has the expression profile a gene in the dataset which would probably have been discarded with a simple variance filter, to illustrate the overall increase in power using ANCOVA. By adjusting for the disease in Figure 3, the strong differential change of expression based on telomere length becomes apparent.

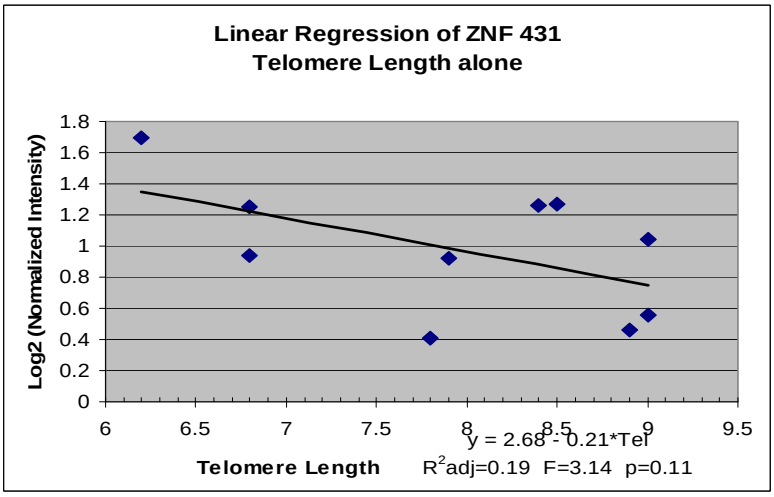

*Figure 2: Linear regression of the RNA expression of the probe for ZNF 431 show s a low adjusted R^2 of 0.19, and an insignificant slope (p=.11). This gene most likely would have been discarded by a simple variance filter. Equation: y=2.68-0.21*Tel R^2_{adj}=0.19 F=3.14 p=0.11*

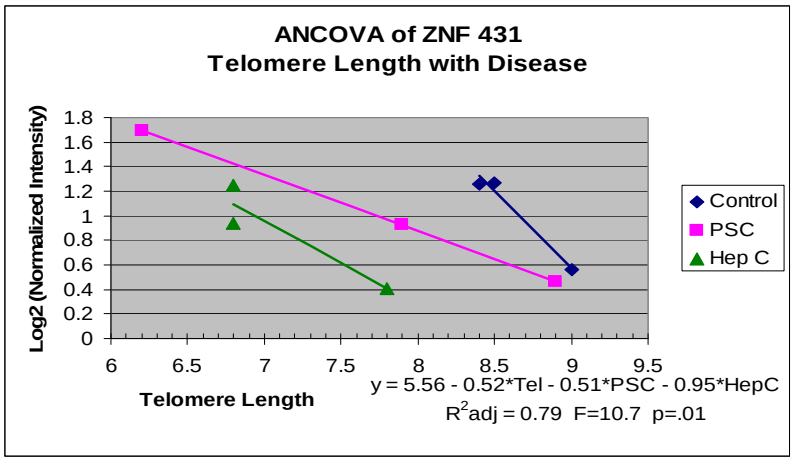

*Figure 3: The data of the same gene are adjusted data for disease using ANCOVA. With this added information, the adjusted R^2 rises to 0.79 with a significant slope (p=.01). This shows that in fact the gene is most likely differentially regulated based on telomere length, and would be included in the subset. Eq'n: y=5.56-0.52*Tel-0.51*PSC-0.95*HepC R^2_{adj}=0.79 F=10.7; p=0.01*

Sample Data Processing

Removal of Bad Data Points

All samples were normalized using the protocol presented by microarray chip manufacturers. Data points which had background contamination, were irregularly shaped, or otherwise unreliable were removed. This represented 3.475% of the data points.

Censored Data

Because ANCOVA relies on a linear model, "left censored" data had to be addressed. The general shape of the fluorescent intensity curve for microarrays is sigmoid. That is, as the sensitivity of the microarray approaches the background noise level, the fluorescent signal does not linearly reflect the mRNA expression level of the sample. Additionally, it is assumed that the majority of human genes will not be expressed at any time, but will still show some signal due to variations in the background noise level. These variations become accentuated when the data is transformed with a log 2 transform as described below.

For this reason, the Negative TrimMean Control, which is the normalized median fluorescent intensity of the negative controls, was used as the baseline for "background noise." All fluorescent intensity readings below this level were artificially raised to it.

Ideally, "right censored" data where the signal becomes saturated should also be addressed. However, because the positive controls (in the form of standardized bacterial RNA spikes) did not provide as clear of a cut point as the Negative TrimMean Control, and because only about 250 genes had signals above the positive control signals that were provided, we deemed it unnecessary to adjust the upper end of the signal.

Log 2 Transform and Duplicate Averaging

The microarray chips were run in duplicates. The data were log 2 transformed, and then averaged afterwards, to reduce the impact of potential outliers. If one of the data points were bad or missing, the other data point was taken as the average instead. If both data points were bad or missing, the data point

was considered missing as well. Replicas which were greater than eight fold different in intensities (that is, that had a log 2 difference greater than 3) were removed as being unreliable.

Repeated measures analysis (in place of averaging the data points first) was deemed to be unnecessary, since we were not interested in internal variations. Additionally, ANCOVA relies on equal variance, which is not a reasonable assumption for microarrays, given the number of potential places in the procedure where an aberrant intensity reading might be introduced. One aberrant intensity reading could grossly inflate the variance, resulting in the entire gene being discarded.

Data Normality

Since ANCOVA depends on normally distributed data, this was checked individually for each gene (each row), using Excel's skewness and kurtosis functions. The data was deemed insufficiently normal and the gene was discarded if one of these values was greater than ± 2*SES or SEK, respectively. The Standard Error of Skewness (SES) was estimated as $\sqrt{6/n}$, while the Standard Error of Kurtosis (SEK) was estimated as $\sqrt{24/n}$ [32].

Removal of Genes with Unknown Function

After the data preprocessing, genes with unknown functions were removed. This included clones, expressed sequence tags (EST), proprietary probes of companies, and others.

ANCOVA Filtering

XL-Stat and ANCOVA Macro

To facilitate a simple ANCOVA analysis of each gene, the program XL-Stat was used (www.XLStat.com). This program by the company Addinsoft (Paris, France) is a supplemental program which runs with Microsoft's Excel software, allowing Excel to handle more advanced statistical analysis, and eliminating the problems associated with transferring data and learning a new syntax for another program. A macro was

written which automatically ran the ANCOVA analysis for each gene. See Appendix B.10 for the code. Data had to be run in sets of no more than three thousand, because of computer RAM memory issues.

Application of ANCOVA for This Dataset

Only genes which had nine or ten data points (up to one missing) were used. Because there were a maximum of ten samples per gene in this dataset, the ANCOVA could not be performed simultaneously against age, telomere length and disease, as would be the normal procedure. Instead, the ANCOVA was run three separate times in pairs: age-disease, age-telomere length and telomere length-disease. The Type I error for falsely rejecting the null hypothesis was set at $\alpha=0.05$ for each pair, giving a total Type I error of 0.147 for at least one false rejection of the null hypothesis, calculated as follows: $P(X\geq1) = 1 - P(X=0) = 1 - (0.95)^3 = 0.147$. Since this is only the preliminary filtering step, additional analysis (hierarchical clustering, etc.) and gene ontology network identification reduces the real Type I error further to a more acceptable level. Because a regression line cannot be created from a single point, the patient with liver cyst disease had to be removed with the age-disease and telomere-length disease adjustments.

Appendix C lists the Excel code used to preprocess much of the data, including the ANCOVA macro code.

MATERIALS AND METHODS

GEO Liver Dataset GSE9588 - Description and Data Processing

Original Study

In May 2008, Schadt et al published a microarray analysis of 427 human liver tissue samples from postmortem and live surgical sources [33]. Their study involved linking microarray expression analysis to single nucleotide polymorphisms (SNiPs), to identify and validate susceptibility genes, and to understand how changes in the networks lead to changes in disease traits, including obesity, diabetes and atherosclerosis. The data were adjusted in the analysis for age, sex and center at which the tissue was processed. The normalized datasets were posted on the National Institutes of Health's Gene Expression Omnibus (GEO) as GSE9588, prior to making adjustments. It was log base ten transformed, and censored to -2 and +2.

Dual Channel Microarray Platform

The microarray platform used is the Rosetta/Merck Human 44k 1.1 platform (GPL4372), which consists of 39,302 60-mer gene probes, of which 1,228 are not publically classified. This dual channel array uses RNA from a control and experimental sample, each labeled with a different dye, that competitive binds to the same microarray probe. The control used was a pooled sample of liver RNA from 191 subjects, labeled with the Cy3 dye channel. The individual subjects were labeled with the Cy5 dye. The samples were processed and analyzed in the protocol previously described [33].

Data Preprocessing

For this analysis, the data was transformed from log 10 to log 2. A simple variance filter captured the 1,965 probes which had a variance of at least 5% of mean standard deviation of the top one hundred genes. This subset was used in all subsequent analyses.

GSE9588 Population Demographics

The dataset was taken from GEO and matched against the population demographics of the subjects. Subjects were selected who were within the age range of 20-29 or 70-81. Subjects whose ages were imputed were removed. Additionally, several subjects (n=11) were removed because of a significant number of missing intensity data points across the genes. The result was 67 Caucasian subjects within these age ranges. Almost a quarter had drug liver risk (medications which affect liver function), because of use of prescription and over the counter medication. One third of the young and one sixth of the old had steatosis (fatty liver), although steatosis was more severe in some of the older subjects. There was one subject in each group with known alcohol risk, although patient data for this category was mostly missing. See Table 2 below.

Table 2: Population Demographics of Subjects Used from GEO Dataset GSE9588.

Age Range #	Gender #	Drug Liver Risk DLR	Steatosis	Alcohol Risk
20-29 years 25	Female 6 Male 19	No 19 = 5 F; 14 M Yes 6 = 1 F; 5 M	Mild 7; 2 w/DLR Moderate 1; Severe 0;	No 8 known Yes 1 known (w/out DLR)
70-81 years 42	Female 24 Male 18	No 34 = 19 F; 15 M Yes 8 = 5 F; 3 M	Mild 4; 1 w/DLR Moderate 1; Severe 2; 2 w/DLR	No 8 known Yes 1 known (w/DLR)

Medication Use

Medication use was not recorded in all cases; none of those in the young group had medications record. In the older group, most were related to heart or possibly stroke conditions, as follows: Coumadin (anticoagulant); Diovan (angiotensin II receptor antagonist for high blood pressure and congestive heart failure); HCTZ (diuretic) and ketoprofen (non-steroid anti-inflammatory, pain killer, reduces fever); Plavix (anticoagulant) and insulin (diabetes); Ecotrin (coated aspirin) and Ziac (diuretic); Coumadin (anticoagulant) and Lasix (diuretic).

MATERIAL AND METHODS

Microarray Analysis

Data preprocessing resulted in a subset of 1,561 genes for the ten liver samples, and 1,965 genes for the GEO liver.

TiGR MeV, DAVID and PANTHER

After data preprocessing, all subsequent microarray analysis for both liver sets were conducted with the TM4 TiGR MultiExperiment Viewer (MeV) 4.0 and the MultiArray Viewer open-source program, available at www.tm4.org/mev.html [34]. Log transformation and variance filtering of the GEO dataset, previously mentioned, were also conducted with TiGR MeV. Gene code conversion and gene ontology clustering were conducted with the National Institutes of Health's DAVID, the Database for Annotation, Visualization, and Integrated Discovery, available at http://david.abcc.ncifcrf.gov [35]. Additional gene ontology clustering and pathway identification was conducted with PANTHER, Protein ANalysis THrough Evolutionary Relationships, available at www.PantherDB.org [36] [37].

Analysis A - Ten Liver Samples

Distance Metric

For Hierarchical Clustering (HCL), Self Organizing Tree Algorithm (SOTA), K-means Clustering, Self Organizing Map (SOM), Cluster Affinity Search Technique (CAST), QT Cluster and others, the distance metric used was Pearson's Correlation.

Pavlidis Template Matching

In addition to the standard clustering methods, the Pavlidis Template Matching was used. Patient age was converted to a ratio between 0 and 1, using the following formula: (patient age - min. age) / (max. age - min. age). The minimum age was 21, the max age was 61. This identified genes whose expression matched this ratio. By subtracting the ratio from 1, genes down regulated with age were identified. A similar

formula was used to identify genes up and down regulated by average telomere length. Minimum telomere length: 6.2. Maximum telomere length: 9.0. Figure 4 shows how this template matching occurs.

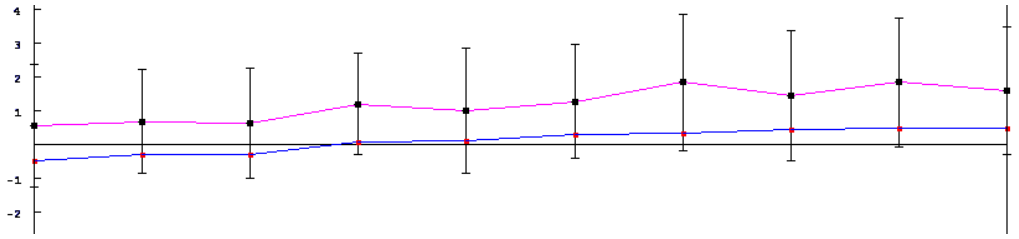

Figure 4: Pavlidis Template Matching by average telomere length, with samples loaded by telomere length. Blue line (bottom): telomere length ratio (0 to 1). Pink line (top): cluster of 83 genes with the closest matching expression pattern. Sample labels on x-axis omitted for space considerations.

Analysis B - GEO Liver Dataset GSE9588

Distance Metric

For Hierarchical Clustering (HCL), Self Organizing Tree Algorithm (SOTA), K-means Clustering, Self Organizing Map (SOM) and others, the distance metric was selected which was best able to cluster the samples by their age group, or which created a gene expression cluster differentiated by the age group. This metric was almost always the Average Dot Product, although the Pearson Correlation metric also came close, and the Pearson.Uncentered metric was used once.

T-test, SAM and ANOVA

In addition to these methods, two group comparisons were made with a t-test and with Significance Analysis for Microarrays (SAM). A two-factor ANOVA was also employed twice, with age against gender or against drug liver risk.

RESULTS

Ten Liver Samples

Hierarchical clustering by several different distance metrics continuously clustered the three patients with Hepatitis C and the three youngest patients, ages 21-29, which partially validates the ANCOVA filtering method as a model. See Fig. 5.

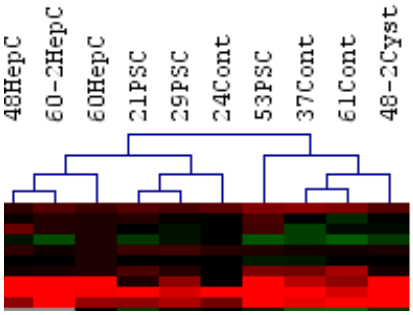

Figure 5: Hierarchical clustering groups the 3 Hepatitis C patients (…HepC) and the 3 youngest patients (21PSC, 24Cont, 29PSC).

Figures A1 - A12 in Appendix A show some of the expression patterns found by various clustering methods.

Differential expression was divided into four categories: up regulated by chronological age, down regulated by age, up regulated with shorter telomere length, down regulated with shorter telomere length. Note: shorter average telomere lengths correspond to an older cellular age.

Out of the subset of 1,561 genes, the following number of genes was found to be differentially regulated:

Age Up: 227 genes

Age Down: 136 genes

Tel. Up: 81 genes

Tel. Down: 159 genes

These genes were loaded into DAVID and PANTHER, using all genes on the microarray as background, to identify pathways, biological processes and molecular functions which had a higher than expected representation. DAVID bases its analysis on an "Enrichment Score," with a higher number being better. Only enrichment scores greater than 1.0 were reported. PANTHER calculates a p-value based on a Chi-square test comparing the observed number of genes to the number expected to appear by random chance.

P-values less than 0.05 were selected. Tables 3and 4 below show details of the results. In DAVID, genes which appear indented in the same table cell are "children" (subsets) of the larger category.

Differentially Regulated Biological Processes and Molecular Functions:

Age Up: Ubiquitin-protein ligase; Golgi apparatus; zinc finger proteins

Age Down: Immunity and defense; receptor; metal cation channels

Telomere Up: None detected.

Telomere Down: Histones; chromatin packaging and remodeling (strong hits: p=E-15 & E-12) and transcriptional regulation.

Table 3: DAVID Functional Analysis Clusters of Ten Liver Samples

Expression	Cluster	Enrichment Score	# genes
Age Up	Ubiquitin	1.46	14
Age Up	Golgi	1.23	6
Age Up	Zinc Finger	1	39
Age Down	Metal Cation Channel	1.01	23
Telomere Up	(None >1.0)		
Telomere Down	Transcription Regulation	1.47	34

Table 4: PANTHER Analysis of Ten Liver Samples.

Analysis	Expression	Description	Ref List	Obs.	Expect	p-value
Molecular Function	Age Up	Ligase	374	13	4.71	3.01E-02
		* Ubiquitin-protein ligase	187	10	2.36	2.48E-02
Biological Process	Age Down	Immunity and defense	1137	20	8.42	8.61E-03
Molecular Function	Age Down	Receptor	1220	22	9.04	2.69E-03
Biological Process	Telomere Down	Nucleoside, nucleotide and nucleic acid metabolism	2732	31	14.95	1.09E-03
		* Chromatin packaging and remodeling	191	16	1.04	1.64E-12
Molecular Function	Telomere Down	Nucleic acid binding	2132	30	11.66	1.87E-05
		* Histone	72	14	0.39	1.08-15

RESULTS

GEO Liver Dataset GSE9588

These analyses identified 165 differentially regulated genes, of which approximately 130 were up regulated in the older group. Genes up regulated by age group consistently clustered by several different methods in an unexpected way. Expression in the young group was generally low. The older group spontaneously clustered into two subgroups with these genes: with and without drug liver risk or steatosis. The genes in the elderly without drug liver risk were up regulated. However, the genes in the elderly with drug liver risk or with severe steatosis (fatty livers) were down regulated, the same as the younger group. These clusters were the only ones which were identifiably clustering based on age group. Figure 6 below shows this unexpected clustering behavior.

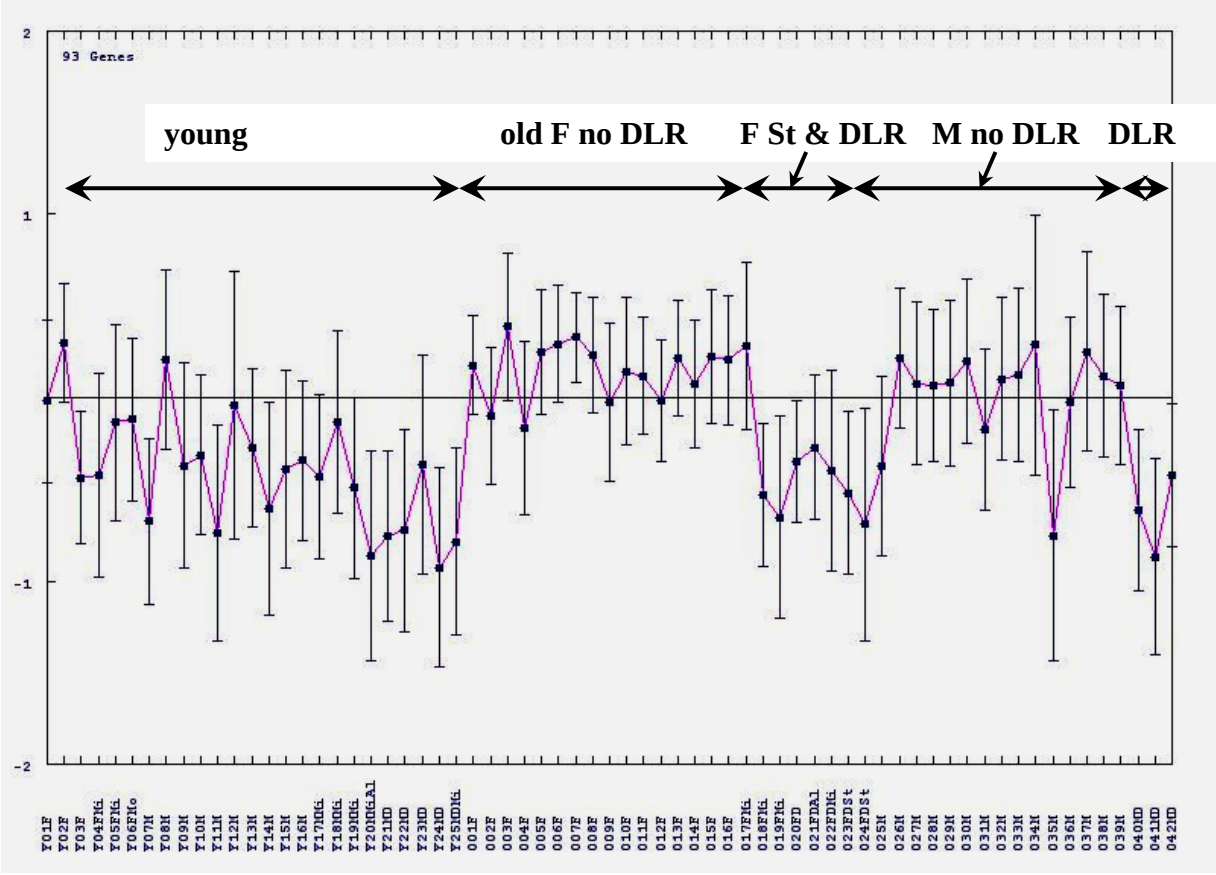

Figure 6: Up regulated. K-means clustering of the sample group. Left side: young (n=25). Right side: older, beginning with the first up regulated subject. The second down regulated group are

older females with mild steatosis (n=2) and drug liver risk (n=5), of which the last two had severe steatosis. Two older males without drug liver risk were also down regulated. The last three subjects are older males with drug liver risk. This pattern was consistent across several different clustering methods.

Clustering also produced about 37 genes with an approximately reversed expression profile, although this pattern was much less consistent and strong. See Figure 7 below.

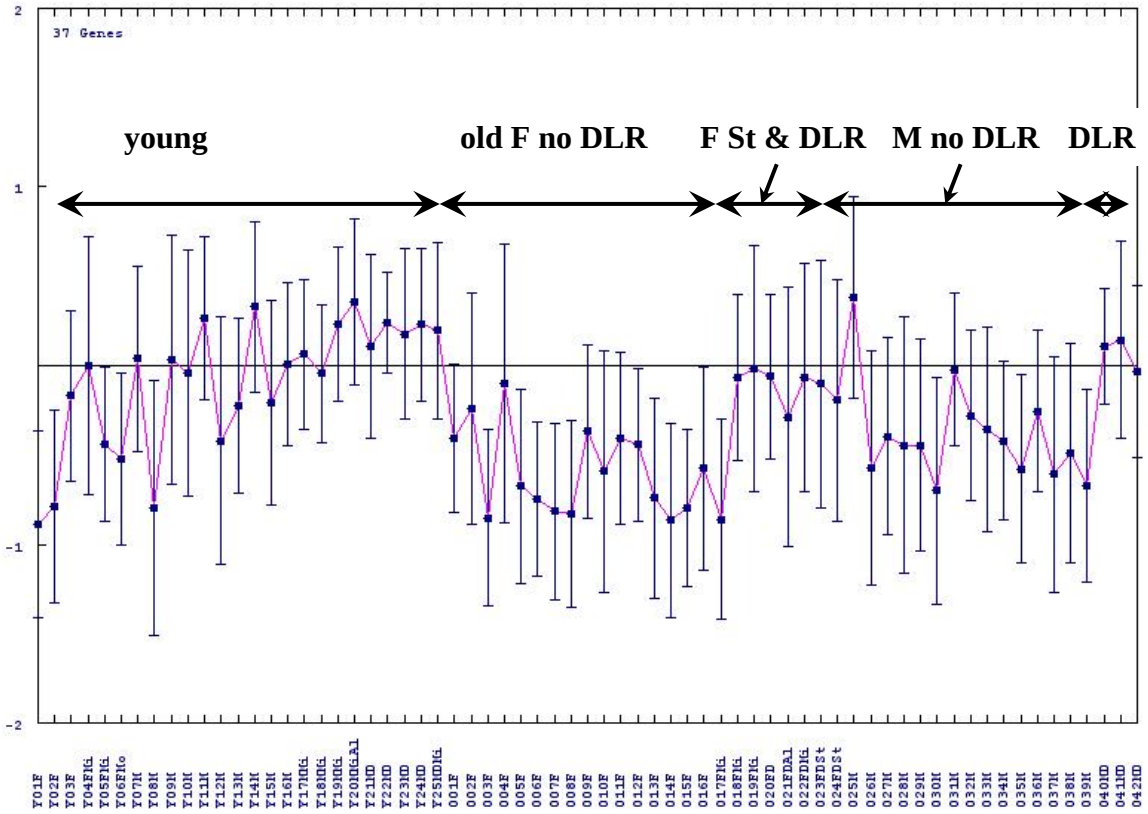

Figure 7: Down regulated. SOTA cluster by the Pearson Uncentered metric resulted in 37 genes which were generally down regulated among older subjects without drug liver risk (right side). The second group near zero are older females with drug liver risk (n=5) with severe steatosis (n=2), and females with mild steatosis (n=2 of 3). The last three subjects are males with drug liver risk. This pattern was less clear cut and less consistent than the opposite regulatory pattern.

Age Up: Nicotine degradation pathway;

Steroid and fatty acid metabolism; carbohydrate metabolism;

Electron transport; oxygenase and reductase; dehydrogenase; cytochrome p450

Glutathione s-transferase.

Mixed: Immunity and defense; chemokines; Interleukin 8.

Age Down: Antibacterial response protein

Table 5: Significantly Regulated Pathways, Biological Processes and Molecular Functions in the GEO Postmortem Liver Set GSE9588, Identified by PANTHER

Type	Expression	Description	Ref List	Obs.	↑↓	Expect	p-value
Pathway	Age Up	Nicotine degradation	9	3	↑ 3	0.05	3.44E-03
Biological Process	Age Up	Lipid, fatty acid and steroid metabolism	699	17	↑15	3.96	1.37E-05
		* Fatty acid metabolism	163	6	↑ 6	0.92	5.15E-02
		* Steroid metabolism	170	15	↑14	0.96	1.04E-11
		- Steroid hormone metabol.	42	5	↑ 5	0.24	9.49E-04
Biological Process	Age Up	Carbohydrate metabolism	534	10	↑ 9	3.03	2.94E-02
		* Other carbohydrate metabolism	53	4	↑ 4	0.3	3.69E-02
Biological Process	Age Up	Electron transport	221	11	↑11	1.25	1.93E-06
Molecular Function	Age Up	Oxidoreductase	550	19	↑19	3.12	8.93E-09
		* Oxygenase	94	12	↑12	0.53	5.97E-11
		* Reductase	157	6	↑ 6	0.89	4.70E-02
Molecular Function	Age Up	Dehydrogenase	205	6	↑ 6	0.84	3.23E-02
Molecular Function	Age Up	Transferase	799	13	↑12	4.53	1.77E-02
		* Other transferase	119	6		0.67	1.06E-02
Biological Process	Age Mixed	Immunity and defense	1188	17	↑12	6.73	1.17E-02
Molecular Function	Age Mixed	Signaling molecule	726	14	↑ 9	4.11	1.87E-03
		* Chemokine	46	4	↑ 2	0.26	2.40E-02
Molecular Function	Age Down	Defense/immunity protein	320	8	↓ **5**	1.81	1.45E-02
		* Antibacterial response protein	34	3	↓ **3**	0.05	3.68E-03

Table 6: Functional Clusters of the GEO Postmortem Liver Set GSE9588 Found by DAVID

Express	#	Cluster name	Enrichment Score	# Genes	Class stringency	Genes
Age Up	1	Cytochrome p450	7.56	14	Low	All similar cytochrome p450
Age Up	2	Reductases	3.23	3	Low	Steroid 5 alpha reductase Cytochrome p450 Aldo-keto reductase
Age Up	3	Glutathione s-transferase	2.83	4	Low	All similar glutathione s-transferases
Age Up	4		2.33	4	Low	pregnancy-zone protein secreted phosphoprotein 2 zymogen granual protein 16 fetuin b
Age Up	5		2.3	3	Low	clusterin-like 1 growth differentiation factor 15 fetuin b
Age All	6	Interleukin 8	2.99	5	Medium	IL8 chemokine (c-c motif) ligand

DISCUSSION

Use of ANCOVA

While using ANCOVA is thought to have improved the overall power of the analysis, it does have limitations associated with assumptions of the test. While some assumptions like normal distribution could be tested directly and automatically, several others could not. In most cases, violations of these assumptions lead to a reduction of power and an omission of a gene that might be of importance. For example, this is true of the assumption of linearity and homogeneity of slopes. With linearity, a non-linear curve is difficult to assess with only three data points per category. Despite these limitations, the overall power and accuracy should be increased, and outweighs some of the genes that may be lost in the process.

Left Censored Data

Additionally, some protocols recommend discarding genes which show no expression. The drawback of this is that it can be difficult to distinguish between a gene with low expression and a gene with no expression. More importantly, a gene which goes from no expression to expression is as important as a gene which goes from low expression to high expression. Raising the expression values to the Negative TrimMean Control, which differs for each chip, eliminates this problem. Analysis of genes which were at this lower level for all samples confirmed that the ANCOVA removed them from the subset of genes used in further analysis.

Genes Up Regulated with Chronological Age

Several gene ontology clusters were up regulated with age. In the ten sample dataset, this includes 15 ubiquitin related genes and six genes of the Golgi apparatus. In the GEO dataset, this includes 37 genes related to cytochrome p450, oxidoreductases and detoxification.

Ubiquitin Enzymes Up Regulated with Age in the 10 Sample Dataset.

In the ten sample dataset, both DAVID and PANTHER found an up regulation of genes related to ubiquitin, which is a protein tag for intracellular translocation. A consecutive series of ubiquitin tags targets a protein for degradation [9]. An up regulation of these enzymes could indicate an increase in the number of proteins that are not folding properly. Another poorly characterized gene was down regulated (SDCCAG10, Entrez ID 10283), which is thought to accelerate protein folding, as was a negative regulator of NFκB and ubiquitin (PPIL5, Entrez ID 122769). Table B1 in Appendix B lists the specific ubiquitin genes that were up regulated, along with their functions.

Cytochrome p450 Up Regulated with Age in GEO Dataset GSE9588

In the GEO dataset, there were a number of biological processes and molecular functions that are all related to cytochrome p450 (CYP): electron transport, oxygenase, reductase, fatty acid metabolism and steroid metabolism. These were highlighted by both DAVID and PANTHER. There was low expression among the young, and among the elderly who had steatosis or drug liver risk. However, expression was high among the elderly without steatosis or drug liver risk. These results were strong and internally very consistent, appearing repeatedly with several different microarray clustering algorithms, usually using the Average Dot Product distance matrix.

CYP in Other Studies

However, the results are unexpected, and appear to conflict with other studies, such as several that show a reduced clearance of CYP3A xenobiotic substrates in the elderly [38], or others which show an overall decrease in surface area of smooth endoplasmic reticulum in rats [39]. There are conflicting results, and variability between subjects tends to be large [40].

Two larger studies looked at CYP expression. Both of these studies investigated actual protein concentrations and enzymatic activity, rather than at mRNA expression levels. The first study of 54 human

subjects [41] found no difference between age groups with cytochrome p450 expression. The second study of 226 human subjects [42] found a significant decrease in CYP concentration among the elderly group.

However, our results correlate very closely with a third study [43]. This particular microarray study is limited because of a very small sample size (two rat and two human livers), and because they compared a one-year old infant to a 78 year old, instead of using a young adult. They found at least a three-fold increase in both CYP and GST genes, including CYP1A1, CYP1A2 and CYP2C18 (not up regulated in this study), as well as GSTA2 and GSTT2. They confirmed mRNA up regulation of CYP1A1 and CYP2C18 using RT-PCR. They tested GST activity using CDNB, and found a small, insignificant *decrease* in GST activity in the 78 year old.

A fourth microarray study of rats [44] found mixed differential expression in thirteen CYP genes, split approximately equally between up and down regulation. This includes a decrease in four CYP3A isoforms (contrasted with an increase here). The study involved 15, 10 and 13 rats at 32, 58 and 84 weeks of age, respectively. None of the specific CYP isoforms differentially expressed were identical to the isoforms found up regulated in this study. They also found mixed differential regulation of Phase II metabolic enzymes, including glutamine s-transferase (GSTA2, GSTA5 and GSTT1).

It is possible that results from this study are biased by an unidentified confounder, such as possibly differences in post mortem versus surgical hepatic biopsies, differences in the center processed, different diseases or cause of death, or differences in care and treatment.

If the up regulation is related to steroid metabolism, it could indicate a feedback mechanism designed to reduce the impact of a chronic presence of steroids, perhaps due to chronic inflammation signals or similar chronic signal. The up regulation of genes in the Golgi complex may be a secondary effect of this change. Since those older patients taking medications were down regulated, this probably does not affect xenobiotic metabolism. It is possible that those people on medications were generally receiving better care, and were in a more "natural" state than those who were not on medications.

Alternatively, this could reflect a reduced regenerative capability in the elderly. A microarray study of transgenic mice which over express hepatocyte growth factor (HGF) had revealed an overall decrease of several CYP genes after partial hepatectomy [45]. The elderly with drug liver risk or steatosis who did not show an increase in CYP gene expression may have had conditions which stressed the body, which may have resulted in HGF release. Several appear to have had heart conditions while others were probably obese. This goes in hand with the decrease in receptors possibly related to regeneration in the ten sample set, discussed below.

Combining these results with those from the ten sample dataset, which showed an increase in ubiquitin, could suggest that there is a problem occurring between transcription and activated enzyme, with the body attempting to compensate by increasing transcription. Similarly, the body may be trying to compensate for the general decrease in hepatic blood flow and volume.

Generally, these results indicate changes in biosynthesis. Recent work has demonstrated a connection between endoplasmic reticulum (ER) stress in the cytoplasm and diabetes, where protein unfolding in response to ER stress is hypothesized to disrupt processes associated with diabetes [33]. The unfolded protein response has also been linked to inflammation and oxidative stress [33], thus offering a relationship between the above discussed molecular changes in metabolism and inflammatory markers.

Receptors Down Regulated by Age in 10 Sample Dataset

Out of the 23 receptors down regulated with age, approximately half were related to the immune system, described below. The remaining eleven receptors were generally not well characterized. However, six of them appear to be related either to development (unspecified, cardiac or mesoderm), cell proliferation or "--genesis" (angiogenesis, neurogenesis, or oncogenesis). It is possible that these receptors are involved in liver regeneration, suggesting either a decreased ability of aged liver to regenerate, or perhaps to

differentiate into hepatocytes and other liver cells. Liver regeneration is thought to decrease with age [39]. Table B2 in Appendix B lists all receptor genes with their known functions.

Immunity and Defense: Down Regulated by Age in Ten Sample Dataset and Mixed Regulation by Age in GEO Dataset GSE9588

Ten Sample Dataset

In the ten sample dataset, PANTHER found these genes down regulated with age. Of the twenty genes, five are cytokines or chemokines, six are expressed in immune cells themselves, with the remaining having other functions. Table B3 lists the genes. Since this dataset had seven people with diseases, the decrease in expression with age reflects a decreased ability to attract immune cells, and a decreased ability of the immune system to respond to diseases presented: Hepatitis C and Primary Sclerosing Cholangitis.

GEO Dataset GSE9588

In the GEO dataset, 17 genes were differentially regulated, but not in a consistent direction. Five of the genes up regulated are related to glutamine s-transferase, which group with cytochrome p450 genes. Of the seven others up related, three are involved in inflammation, which has been shown to increase with age in a recent microarray rat study [46]. For the five genes down regulated, three are chemotaxins, which matches results from the ten sample dataset. The other two are apolipoproteins. Table B4 lists the genes and their specific functions.

Histones and Transcription Down Regulated with Shorter Telomeres - Ten Sample Dataset

In the ten sample dataset, PANTHER found strong down regulation of several Histone H4 genes, and chromatin packaging and remodeling. In addition to the histones themselves, there were several ribosomal proteins. The remaining genes were principally related to transcription regulation. Table B6 lists the specific genes that were down regulated, along with their functions. As telomeres become shorter, cells go into replicative senescence to exit the cell cycle. When they do, chromatin condenses in particular in areas usually used for replication [47], in chromatin areas described as senescence-associated heterochromatic

foci (SAHF) [48]. While biological conclusions can be drawn with the results from the PANTHER tool, it is unknown why DAVID had no hit using the same dataset. Since ribosomes are part of the ER, down regulation may indicate less resident ER protein turnover/biogenesis, increasing the likelihood for dysfunction. It is unclear why only Histone H4 would be found to be down regulated, since this subunit normally works in conjunction with Histone H2A, H2B and H3 proteins in the nucleosome [9].

Genes Up Regulated with Shorter Telomere Length - Ten Sample Dataset

Although microarray clustering found more than 120 genes up regulated with shorter telomere length, neither DAVID nor PANTHER were able to find genes within that group that had similar gene ontologies or networks. This could either indicate a large number of false positives, or a network system more complicated than either program could decipher.

CONCLUSION AND HYPOTHESES

Conclusion

We have shown that by using specific statistical techniques we can identify a molecular transcription profile in old and diseased liver specimen that are related to shortened telomeres, thus linking a phenotypically accessible marker in tissues with specific molecular events. It is very much debated whether telomere shortening contributes to the aging process, since rodents, for example, have much longer telomeres than humans that do not go beyond a critical length in old age [9]. Observations of molecular changes as they occur with telomere shortening as shown here, even accelerated with disease as reported previously [28], confirm the critical role of telomeres as these changes go along with a general architectural change in liver, including alterations in the cellular composition and inflammatory infiltration [28]. However, observational studies cannot decipher cause and effect and the sequence of events, but rather dedicated experiments will be needed.

In summary, these microarray studies found an increase in ubiquitin and cytochrome p450 related genes with increasing chronological age. Some receptors had decreased with chronological age. Immunity and defense genes showed mixed differential expression with age. Shorter telomere length was associated with a decrease in expression of several Histone H4 genes, as well as other genes related to transcription regulation. A pattern in genes up regulated with shorter telomere length could not be found.

Microarrays have proven extremely useful as a hypothesis generating tool. Here, we present seven hypotheses from the results found in this study.

CONCLUSION AND HYPOTHESES

Hypotheses

Hypothesis 1: Six receptors (below) are linked to liver regeneration.

Reasoning: Six receptors apparently involved in some aspect of development were down regulated with age in the ten sample set. All but ENG are poorly characterized, at least with regard to liver expression. Since liver regeneration decreases with age, these receptors may play a role in liver regeneration, with cell proliferation or differentiation. See also Table B2 in Appendix B.

Entrez ID	Gene Symbol	Gene type	Known or Suspected Function
2022	ENG	Endoglin	Angiogenesis via TGF-β
2041	EPHA1	Ephrin receptor	Neurogenesis
10154	PLXNC1	Plexin	Tyr. kinase; IL6 & IL8 secretion
169611	OLFML2A	Olfactomedin-like	Neurogenesis
64123	ELTD1	EGF latrophilin 7 transmemb.	G-protein coupled receptor; cardiac development
10234	LRRC17	Leucine Rich Repeat	Development processes

Hypothesis 2a: Elderly are less able to respond effectively to liver diseases than the young.

Hypothesis 2b: Elderly are more prone to chronic inflammation of the liver than the young.

Reasoning: Decrease in immunity and defense genes in elderly subjects with diseases. Increase in immunity and defense genes related to inflammation.

Hypothesis 3: Elderly have more problems with protein misfolding in the liver than the young.

Reasoning: Increase in ubiquitin systems in the elderly, along with a decrease in transcription with shorter telomeres.

Telomere Hypotheses:

Hypothesis 4: Senescent hepatocytes have lower overall rates of transcription than non-senescent hepatocytes.

Reasoning: Shorter telomeres correspond to senescent hepatocytes. Decreased expression of ribosomes and genes related to transcription in those with short telomeres.

Hypothesis 5a: Chromatin in the nucleus of senescent hepatocytes is more condensed in senescent cells.

Hypothesis 5b: A portion of the condensed chromatin does not require the presence of Histone H4.

Reasoning: Decrease in several Histone H4 proteins in those with short telomeres.

Cytochrome p450 Hypotheses:

Hypothesis 6a: CYP mRNA levels do not correlate linearly with CYP activated protein levels in the elder; mRNA levels are higher than CYP protein concentrations would predict.

Hypothesis 6b: CYP activated protein levels are low in the elderly because of decreased mRNA stability, problems with protein folding, decreased CYP stability or increased inactivation of CYP.

Hypothesis 6c: A feedback mechanism induces CYP and GST transcription in the elderly, in an attempt to compensate for low CYP activated protein levels and/or low hepatic blood flow.

Reasoning: High CYP mRNA levels in the elderly that were found in this study do not seem to correlate with other studies which have found no change or decreased activated CYP protein in the elderly. Hepatic blood flow is generally lower in the elderly. Also, ubiquitin enzymes were found to be up regulated in this study.

Hypothesis 7a: Hepatocyte turnover is higher in elderly with steatosis or drug liver risk than elderly without steatosis or drug liver risk.

Hypothesis 7b: Hepatocyte Growth Factor (HGF) levels are higher in elderly with steatosis or drug liver risk than they are in elderly without steatosis or drug liver risk.

Reasoning: HGF suppresses CYP expression. CYP is higher in elderly without steatosis or drug liver risk than in elderly with steatosis or drug liver risk.

These hypotheses could be tested with a variety of molecular biology techniques that are currently available, such as proteomics and immunohistochemistry. Some, such as the CYP hypotheses, would only require measuring mRNA, protein levels and protein activity from the same liver. Others, such as identifying the feedback mechanism, would require more a more elaborate experimental design. The enduring problem with this, of course, is the difficulty of obtaining a human liver biopsy, since it is not ethical to obtain these from healthy volunteers, due to the risk of infection.

References

1. Finch, C.E., *Longevity, senescence, and the genome*. 1990, Chicago: University of Chicago Press. xv, 922 p.

2. Berner, Y.N. and F. Stern, *Energy restriction controls aging through neuroendocrine signal transduction.* Ageing Res Rev, 2004. **3**(2): p. 189-98.

3. Zeuthen, E., *Oxygen uptake as related to body size in organisms.* Q Rev Biol, 1953. **28**(1): p. 1-12.

4. Harman, D., *Aging and disease: extending functional life span.* Ann N Y Acad Sci, 1996. **786**: p. 321-36.

5. Sastre, J., F.V. Pallardo, and J. Vina, *The role of mitochondrial oxidative stress in aging.* Free Radic Biol Med, 2003. **35**(1): p. 1-8.

6. Bordone, L. and L. Guarente, *Calorie restriction, SIRT1 and metabolism: understanding longevity.* Nat Rev Mol Cell Biol, 2005. **6**(4): p. 298-305.

7. van der Vlies, D., et al., *Endoplasmic reticulum resident proteins of normal human dermal fibroblasts are the major targets for oxidative stress induced by hydrogen peroxide.* Biochem J, 2002. **366**(Pt 3): p. 825-30.

8. Hayflick, L., *The longevity of cultured human cells.* J Am Geriatr Soc, 1974. **22**(1): p. 1-12.

9. Alberts, B., *Molecular biology of the cell*. 4th ed. 2002, New York: Garland Science. xxxiv, 1463, [86] p.

10. Yung, R.L., *Changes in immune function with age.* Rheum Dis Clin North Am, 2000. **26**(3): p. 455-73.

11. Licastro, F., et al., *Innate immunity and inflammation in ageing: a key for understanding age-related diseases.* Immun Ageing, 2005. **2**: p. 8.

12. Sarkar, D. and P.B. Fisher, *Molecular mechanisms of aging-associated inflammation.* Cancer Lett, 2006. **236**(1): p. 13-23.

13. Salminen, A., et al., *Activation of innate immunity system during aging: NF-kB signaling is the molecular culprit of inflamm-aging.* Ageing Res Rev, 2007.

14. Franceschi, C., et al., *Inflammaging and anti-inflammaging: a systemic perspective on aging and longevity emerged from studies in humans.* Mech Ageing Dev, 2007. **128**(1): p. 92-105.

15. Krtolica, A. and J. Campisi, *Cancer and aging: a model for the cancer promoting effects of the aging stroma.* Int J Biochem Cell Biol, 2002. **34**(11): p. 1401-14.

16. Mariotti, M., et al., *Interleukin 1 alpha is a marker of endothelial cellular senescent.* Immun Ageing, 2006. **3**: p. 4.

17. Shelton, D.N., et al., *Microarray analysis of replicative senescence.* Curr Biol, 1999. **9**(17): p. 939-45.

18. Zhang, H., K.H. Pan, and S.N. Cohen, *Senescence-specific gene expression fingerprints reveal cell-type-dependent physical clustering of up-regulated chromosomal loci.* Proc Natl Acad Sci U S A, 2003. **100**(6): p. 3251-6.

19. Marchesini, G., et al., *Galactose elimination capacity and liver volume in aging man.* Hepatology, 1988. **8**(5): p. 1079-83.

20. Wynne, H.A., et al., *The effect of age upon liver volume and apparent liver blood flow in healthy man.* Hepatology, 1989. **9**(2): p. 297-301.

21. Ramanujan, V.K. and B.A. Herman, *Aging process modulates nonlinear dynamics in liver cell metabolism.* J Biol Chem, 2007. **282**(26): p. 19217-26.

22. Serste, T. and N. Bourgeois, *Ageing and the liver.* Acta Gastroenterol Belg, 2006. **69**(3): p. 296-8.

23. Barnett, Y.A. and C.R. Barnett, *Aging methods and protocols*. Methods in molecular medicine ;. 2000, Totowa, N.J.: Humana Press. xiv, 385 p.

24. Wauthier, V., R.K. Verbeeck, and P.B. Calderon, *The effect of ageing on cytochrome p450 enzymes: consequences for drug biotransformation in the elderly.* Curr Med Chem, 2007. **14**(7): p. 745-57.

25. *Liver Disease Information Center*. 2007 September 27 [cited 2008 September 2]; Available from: http://www.liverfoundation.org/education/info.

26. Floreani, A., *Liver diseases in the elderly: an update.* Dig Dis, 2007. **25**(2): p. 138-43.

27. Bellentani, S., et al., *Prevalence of and risk factors for hepatic steatosis in Northern Italy.* Ann Intern Med, 2000. **132**(2): p. 112-7.

28. Wiemann, S.U., et al., *Hepatocyte telomere shortening and senescence are general markers of human liver cirrhosis.* Faseb J, 2002. **16**(9): p. 935-42.

29. Young, M.B., et al., *Analysis of gene expression in carbon tetrachloride-treated rat livers using a novel bioarray technology.* Pharmacogenomics J, 2003. **3**(1): p. 41-52.

30. Canales, R.D., et al., *Evaluation of DNA microarray results with quantitative gene expression platforms.* Nat Biotechnol, 2006. **24**(9): p. 1115-22.

31. Tadesse, M.G., J.G. Ibrahim, and G.L. Mutter, *Identification of differentially expressed genes in high-density oligonucleotide arrays accounting for the quantification limits of the technology.* Biometrics, 2003. **59**(3): p. 542-54.

32. Brown, J., *Skewness and Kurtosis*, in *JALT Testing and Evaluation SIG Newsletter*. 1997. p. 18-20.

33. Schadt, E.E., et al., *Mapping the genetic architecture of gene expression in human liver.* PLoS Biol, 2008. **6**(5): p. e107.

34. Saeed, A., et al., *TM4: a free, open-source system for microarray data management and analysis.* Biotechniques, 2003. **34**(2): p. 374-378.

35. Dennis, G., Jr., et al., *DAVID: Database for Annotation, Visualization, and Integrated Discovery.* Genome Biol, 2003. **4**(5): p. P3.

36. Thomas, P.D., et al., *PANTHER: a library of protein families and subfamilies indexed by function.* Genome Res, 2003. **13**(9): p. 2129-41.

37. Thomas, P.D., et al., *Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W645-50.

38. Cotreau, M.M., L.L. von Moltke, and D.J. Greenblatt, *The influence of age and sex on the clearance of cytochrome P450 3A substrates.* Clin Pharmacokinet, 2005. **44**(1): p. 33-60.

39. Schmucker, D.L., *Age-related changes in liver structure and function: Implications for disease?* Exp Gerontol, 2005. **40**(8-9): p. 650-9.

40. Parkinson, A., et al., *The effects of gender, age, ethnicity, and liver cirrhosis on cytochrome P450 enzyme activity in human liver microsomes and inducibility in cultured human hepatocytes.* Toxicol Appl Pharmacol, 2004. **199**(3): p. 193-209.

41. Schmucker, D.L., et al., *Effects of age and gender on in vitro properties of human liver microsomal monooxygenases.* Clin Pharmacol Ther, 1990. **48**(4): p. 365-74.

42. Sotaniemi, E.A., et al., *Age and cytochrome P450-linked drug metabolism in humans: an analysis of 226 subjects with equal histopathologic conditions.* Clin Pharmacol Ther, 1997. **61**(3): p. 331-9.

43. Thomas, R.P., et al., *Age-associated changes in gene expression patterns in the liver.* J Gastrointest Surg, 2002. **6**(3): p. 445-53; discussion 454.

44. Mori, K., et al., *Hepatic transcript levels for genes coding for enzymes associated with xenobiotic metabolism are altered with age.* Toxicol Pathol, 2007. **35**(2): p. 242-51.

45. Kakizaki, S., et al., *Gene expression profiles of drug-metabolizing enzymes and transporters with an overexpression of hepatocyte growth factor.* Liver Int, 2007. **27**(1): p. 109-19.

46. Singh, P., et al., *Lymphoid neogenesis and immune infiltration in aged liver.* Hepatology, 2008. **47**(5): p. 1680-90.

47. Adams, P.D., *Remodeling of chromatin structure in senescent cells and its potential impact on tumor suppression and aging.* Gene, 2007. **397**(1-2): p. 84-93.

48. Narita, M., et al., *Rb-mediated heterochromatin formation and silencing of E2F target genes during cellular senescence.* Cell, 2003. **113**(6): p. 703-16.

Appendix A - Cluster Expression Profiles

Ten Liver Samples - Age Up

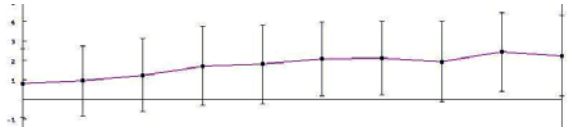

Fig. A1: CAST Cluster, age up, 13 genes

Ten Liver Samples - Age Down

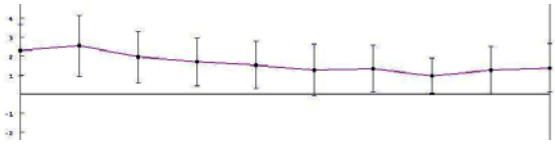

Fig. A2: CAST Cluster, age down, 43 genes

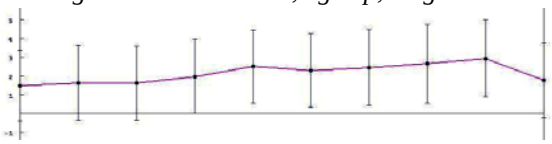

Fig A3: KMC, age up, 173 genes.

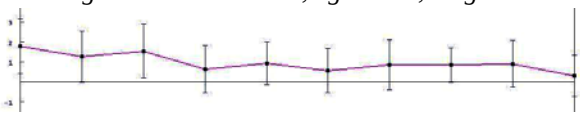

Fig. A4 KMC age down, 98 genes

Fig. A5: QTC, age up, 45 genes

Fig A6: QTC, age down, 47 genes

Fig A7: PTM, age up, 95 genes
Top line: actual cluster;
Bottom: age ratio

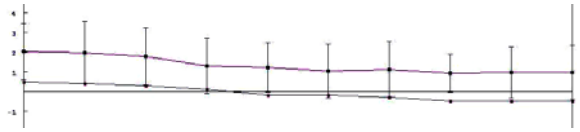

Fig A8: PTM, age down, 81 genes
Top line: actual cluster;
Bottom: reversed age ratio

Ten Liver Smpl - Telomere Long-Short Up

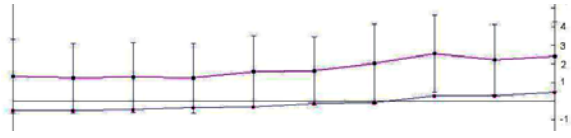

Fig A9: PTM Tel. Long-Short Up, 120 genes
Top line: actual cluster; Bottom: rev. tel. ratio

Ten Liver Smpl - Telomere Long-Short Down

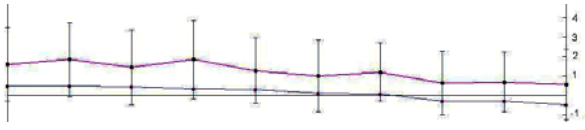

Fig A10: PTM Tel. Long-Short Down 83 genes
Top line: actual cluster; Bottom: telomere ratio

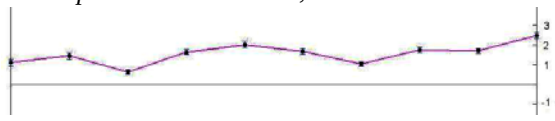

Fig. A11: QTC Telomere Long-Short Up #1, 37 genes

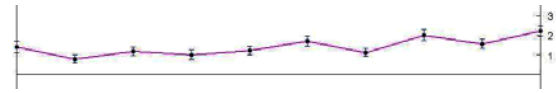

Fig. A12: QTC Telomere Long-Short Up #2, 26 genes

Appendix B Tables of Gene Regulation

List of specific genes up or down regulated by age or telomere length for each dataset. Telomere expression direction refers to regulation from longest to shortest (i.e. by increasing cellular age).

Table	Category Subset	Type	Age/Tel.	Expression	Dataset
B1	Ubiquitin ligase	Function	Age	Up	10 smpl
B2	Receptor	Function	Age	Down	10 smpl
B3	Immunity & Defense	Function	Age	Down	10 smpl
B4	Immunity & Defense	Function	Age	Mixed	GSE9588
B5.1	Cytochrome p450	Biological Process	Age	Up	GSE9588
B5.2	Cytochrome p450	Molecular Function	Age	Up	GSE9588
B5.3	Cytochrome p450	Substrates & Function	Age	Up	GSE9588
B6	Transcription Reg.	Function	Telomere	Down	10 smpl

Table B1: Genes Up Regulated by Age - Ubiquitin Ligases - 10 Sample Set

Entrez Gene ID	Gene Symbol	Gene Type	Ubiquitin protein ligase or ligase (parent)	Liver Expression	Organ of major expression	Cellular Location	Function
6477	SIAH1	Seven in ab-sentia homol.	Ubi-Ligase	Lo	Down reg. in HCC	Cyt; nucl (lo)	Ubiquitin & degradation of proteins
8924	HERC2	Hect domain & RLD 2	Ubi-Ligase				E3 ubiquitin ligase?
9690	UBE3C	Ubiquitin protein ligase E3	Ubi-Ligase		Skeletal muscles	Nucl	E3 ubiquitin ligase
23077	MYCBP2	MYC binding protein	Ubi-Ligase				Tuberin ubiquination and degradation through PAM?
26054	SENP6	SUMO1/sentri n specific peptidase	Ubi-Ligase		Testis, ov-ary, prost.	Nucl	Activation of SUMO (ubiquitin-like molec.)
57154	SMURF1	SMAD specific E3 ubi. pro. ligase	Ubi-Ligase				For SMADs specific to bone morphogenic protein (BMP) pathway
51320	RKHD2	ring finger and KH domain	Ubi-Ligase				RNA binding. Posttranscriptional regulation?
83737	ITCH	Itchy homolog E3 ubi. protein ligase	Ubi-Ligase	Y		Cyt; nucl	Regulates several transcription factors; immune response?
92979	MARCH9	Membrane-asso-ciated ring finger	Ubi-Ligase				Sorting of MHC-I, CD4 and ICAM1 to lysosomes?
7726	TRIM26	Tripartite motif	Ubi-Ligase			Cyt	Unknown; DNA binding? MHC I?
2180	ACSL1	Acyl-CoA sythetase	Ligase	Hi		Mit, ER, Peroxis.	Activation of fatty acids for lipid synthesis or fatty acid degradation
23305	ACSL6	Acyl-CoA sythetase	Ligase		Brain; hemato-poetic cells	Mit, ER, peroxis.	Activation of fatty acids for lipid synthesis or fatty acid degradation
3376	IARS	Isoleucine-tRNA synthetase	Ligase				Links isoleucine tRNA to its amino acid
10283	SDCCAG10	Colon Cancer antigen	Im-mun & Def.	*Down* regulated. Not well characterized.			Accelerates protein folding.
122769	PPIL5	Peptidylprolyl isomerase	Liver	*Down* regulated.			Neg. regulator of NFkB; ubiquitin

Table B2: Genes Down Regulated by Age - Receptors - Ten Sample Dataset [35, 36]

Entrez Gene ID	Gene Symbol	Gene Type	Organ or Cell known to express	Cellular Location	Function or Ligand	Immunity and Defense	Development or --genesis?
1370	CPN2	Carboxypeptidase		Secreted		Y	
1436	CSF1R	Colony stimulating factor receptor		Membrane	CST controls macrophages	Y	
3598	IL13RA2	IL 13 receptor		Membrane	Internalizes IL13? Not a signal mediator	Y	
3820	KLRB1	Killer cell lectin-like receptor	Natural killer cell	Membrane	Regulation of NK cell function?	Y	
3822	KLRC2	Killer cell lectin-like receptor	Natural killer cell	Membrane	MHC Class I molecule	Y	
3823	KLRC3	Killer cell lectin-like receptor	Natural killer cell	Membrane	MHC Class I molecule	Y	
26253	CLEC4E	C-type lectin domain	Peritoneal macrophages	Membrane	Immune surveillance processes? Inflammation role?	Y	
50856	CLEC4A	C-type lectin domain	Dendritic cells	Membrane	Inflammatory and immune response.	Y	
83953	FCAMR	Fc receptor for IgA, IgM			Unknown	Y	
51338	MS4A4A	Membrane spanning 4-domains	Hematopoietic cell lines	Membrane	Signal transduction? Pseudogene?	Y	
64231	MS4A6A	Membrane spanning 4-domains	B-cell, monocyte cell lines	Membrane	Unknown	Y	
2022	ENG	Endoglin	Vascular endothelial cells	Membrane	TGF-beta. Binding of endothelial cells. Angiogenesis, oncogenesis.	N	Y
2041	EPHA1	EPH receptor	Over expressed in carcinomas	Membrane	Ephrin. Neurogenesis, mesoderm devel., cell prolif.	N	Y
10154	PLXNC1	Plexin C1		Membrane	Viral receptor; endogenous unknown, devel. processes.	N	Y
169611	OLFML2A	Olfactomedin-like		Membrane	Unknown; neurogenesis	N	Y
64123	ELTD1	EGF, latrophilin transmembrane	Smooth cells	Membrane	Cardiac development?	N	Y
10234	LRRC17	Leucine rich repeat	Liver-low		Unknown; extracellular matrix. Development processes.	N	Y
203190	LGI3	Leucine rich repeat	High in brain and lung	Secreted	Unknown; extracellular matrix	N	N
219527	LRRC55	Leucine rich repeat			Unknown; extracellular matrix	N	N
2893	GRIA4	Glutamate receptor	Nerve cells	Membrane	Glutamate neurotransmission	N	N
9936	CD302	CD 302 molecule			Endocytosis	N	N
55697	VAC14	Vac 14 homolog			Unknown	N	N

Table B3: Genes Down Regulated by Age in Ten Sample Set - Immunity & Defense Genes [35] [36]

Entrez Gene ID	Gene Symbol	Gene Type	Organ or cell expression	Cellular Location	Function
1370	CPN2	Carboxypeptidase		Secrete	Unknown
1436	CSF1R	Colony stimulating factor receptor		Membrane	Receptor for CST1, cytokine controls production, differentiation and function of macrophages
2162	F13A1	Coagulation Factor XIII	Liver; plasma	Cytoplasm; Secreted	Coagulation cascade
3134	HLA-F	MHC Class I		ER, Golgi	Major Histocompatibility Complex
3433	IFIT2	Interferon-induced protein		Unknown	Unknown
3598	IL13RA2	Interleukin 13 receptor		Membrane	Internalizes IL13? Not a signal mediator
3603	IL16	Interleukin 16		Secreted	Cytokine. Chemoattractant of CD4+ lymphocytes
3820	KLRB1	Killer cell receptor	NK cells	Membrane	Regulation of NK cell function?
3822	KLRC2	Killer cell receptor	NK cells	Membrane	
3823	KLRC3	Killer cell receptor	NK cells	Membrane	
5366	CCL21	Chemokine (C-C motif)	T-cell lines		Chemokine
6368	CCL23	Chemokine (C-C motif)	Liver-high	Secreted	Cytokine. Chemoattractant
7980	TFPI2	Tissue factor pathway inhibitor	Liver-high	Secreted	Blood clotting. Plasmin inhibitor.
9516	LITAF	TNF factor	Liver; spleen		NFkB transcriptional regulation of a chemokine
10283	SDCCAG10	Colon Cancer antigen			Accelerates protein folding.
26253	CLEC4E	C-type lectin domain		Membrane	Immune surveillance processes? Inflammation role?
50856	CLEC4A	C-type lectin domain	Dendritic cells	Membrane	Inflammatory and immune response.
55303	GIMAP4	GTPase IMAP family member	T- & B-cells		GTPase activity
115908	CTHRC1	Collagen triple helix repeat	Chondrocyte-like cells	Secreted	Neg. regulator of collagen matrix
122769	PPIL5	Peptidylprolyl isomerase	Liver		Neg. regulator of NFkB; ubiquitin

Table B4: Genes Up or Down Regulated by Age - GSE9588 Dataset - Immunity & Defense Subcategory

Entrez ID	Gene Symbol	Gene Type	↑↓	CYP related?	Cellular Location	Function or ligand
1401	CRP	C-reactive protein	↓	N	Secreted	Antibacterial; cytokine; DNA scavenger
6288	SAA1	Serum amyloid	↓	N	Secreted (plasma)	Apolipoprotein of HDL complex
6289	SAA2	Serum amyloid	↓	N		Dysfunctional form of SAA1; related to obesity and inflammation
6374	CXCL5	Chemokine (C-X-C motif) ligand	↓	N	Secreted	Chemotaxin for neutrophil activation; inflammation
6364	CCL20	Chemokine (C-C motif) ligand	↓	N	Secreted	Chemotaxin for lymphocytes
9650	CCL4L1	Chemokine (C-C motif) ligand	↑	N		Inflammation mediated by chemokine
3576	IL8	Interleukin 8	↑	N	Secreted	Inflammation mediated by chemokine
2205	FCER1A	Fc fragment of IgE, high affinity	↑	N	Membrane	IgE receptor involved in allergy response.
2353	FOS	v-fos osteosarcoma oncogene	↑	N	Nucleus	Transcription factor with JUN/AP-1
3164	NR4A1	Nuclear receptor	↑	N	Nucleus	Orphan nuclear receptor induced by growth hormones; translocation to mitochondria induces apoptosis
5858	PZP	Pregnancy-zone protein	↑	N		Blood coagulation; serine protease inhibitor
51129	ANGPTL4	Angiopoietin-like	↑	N		Angiopoeitin
2938	GSTA1	Glutathione s-transferase	↑	Y		Detoxification; antioxidant
2939	GSTA2	Glutathione s-transferase	↑	Y		Detoxification; antioxidant
2952	GSTT1	Glutathione s-transferase	↑	Y		Detoxification; antioxidant
221357	GSTA5	Glutathione s-transferase	↑	Y		Detoxification; antioxidant
5475	PPEF1	Protein phosphatase, EF-hand Ca binding	↑	Y		Regulation of carbohydrate metabolism

Table B5.1: Genes Up Regulated by Age - GEO GSE9588 Dataset - Biological Processes [36]

← Entrez Gene ID	Gene Symbol	Gene Type	DAVID or PANTHER	Nicotine Degrad.	Electron Transport	Lipid, Fatty acid, Steroid Metabol.	Steroid Metabolism SM	Steroid Hormone Metabolism	Fatty acid Metabolism	Carbohydrate Metabolism CM	Other Carbo-hydrate Metabol.	Detoxification
Parent							LFSM	SM	LFSM		CM	ID
p-value				.001	E-08	E-06	E-12	E-04	.008	.01	.01	.002
Obs **RefList**				3 9	11 221	15 699	14 170	5 42	6 163	9 534	4 53	4 63
124 **125** **126**	ADH1A ADH1B ADH1C	Alcohol dehydrogenase	P							Y	Y	
127	ADH4	Alcohol dehydro.	P							Y	Y	
155	ADRB3	Andrenergic rcpt	P									
635	BHMT	Methyltransferase	P									
1543	CYP1A1	Cytochrome p450	DP		Y	Y	Y		Y			
1548 **1549** **1553**	CYP2A6 CYP2A7 CYP2A13	Cytochrome p450	DP	Y	Y	Y	Y		Y			
1555 **1556**	CYP2B6 CYP2B7	Cytochrome p450 CYP pseudogene	DP D		Y --	Y --	Y --	Y --				
1558	CYP2C8	Cytochrome p450	DP		Y	Y	Y		Y			
1576 **1551**	CYP3A4 CYP3A7	Cytochrome p450	DP		Y	Y	Y	Y				
1581 **1592** **1595**	CYP7A1 CYP26A1 CYP51A1	Cytochrome p450	DP		Y	Y	Y					
2938 **2939** **221357**	GSTA1 GSTA2 GSTA5	Glutathione S-transferase	DP									Y
2952	GSTT1	" "	DP									Y
5105	PCK1	Carboxykinase	P							Y		
5475	PPEF1	Phosphatase, Ca	P							Y		
6519	SLC3A1	a.a. transporter	P							Y		
6716	SRD5A2	Steroid reductase	P									
6718	AKR1D1	Aldoketo reductase	P									
6783 **6822**	SULT1E1 SULT2A1	Sulfotransferase	P			Y	Y	Y				
6898	TAT	T NH_3 transferase	P									
7364	UGT2B7	Glucuronosyltransf	P			Y	Y	Y		Y		
8424	BBOX1	Dioxygenase	P									
10891	PPARGC1A	Peroxisome recep.	P			Y						
134147	CMBL	Carboxymethylene-butenolidase hom	P							Y		

Table B5.2: Genes Up Regulated by Age - GEO GSE9588 Dataset - Molecular Function [36]

← Entrez Gene ID	Gene Symbol	Gene Type	Oxidoreductase	Oxygenase	Reductase	Dehydrogenase	Transferase	Other transferase
Parent				←OR	OR	OR		←T
p-value			E-11	E-12	.0076	.032	.0027	.0016
Obs **Ref List**			19 550	12 94	6 157	6 205	12 799	6 119
124 **125** **126**	ADH1A ADH1B ADH1C	Alcohol dehydrogenase	Y		Y	Y		
127	ADH4	" "	Y		Y	Y		
155	ADRB3	Andrenergic rcptr						
635	BHMT	Methyltransferase					Y	
1543	CYP1A1	Cytochrome p450	Y	Y				
1548 **1549** **1553**	CYP2A6 CYP2A7 CYP2A13	Cytochrome p450	Y	Y				
1555 **1556**	CYP2B6 CYP2B7	Cytochrome p450	Y --	Y --				
1558	CYP2C8	Cytochrome p450	Y	Y				
1576 **1551**	CYP3A4 CYP3A7	Cytochrome p450	Y	Y				
1581	CYP7A1	Cytochrome p450	Y	Y				
1592	CYP26A1	Cytochrome P450	Y	Y				
1595	CYP51A1	Cytochrome P450	Y	Y			Y	
2938 **2939** **221357**	GSTA1 GSTA2 GSTA5	Glutathione S-transferase					Y Y --	Y Y --
2952	GSTT1	" "					Y	Y
5105	PCK1	Carboxykinase						
5475	PPEF1	Phosphatase, Ca						
6519	SLC3A1	a.a. transporter						
6716	SRD5A2	Steroid reductase	Y			Y		
6718	AKR1D1	Aldoketo reductase	Y		Y			
6783 **6822**	SULT1E1 SULT2A1	Sulfotransferase					Y	Y
6898	TAT	T NH_3 transferase					Y	
7364	UGT2B7	Glucuronosyltransf.					Y	
8424	BBOX1	Dioxygenase	Y	Y				
10891	PPARGC1A	Peroxisome recep.						
29947	DNMT3L	Methyltransferase					Y	
134147	CMBL	Carboxymethylene-butenolidase hom.						

Table B5.3: Genes Up Regulated by Age - GSE9588 Dataset - Substrates & Functions [35]

↓ Entrez Gene ID	Gene Symbol	Gene Type	Cellular Location	Normal Liver Expression?	Mitochondrial?	Reactive Oxidative Species (ROS)?	Obesity?	Some substrates or Functions
124 **125** **126**	ADH1A ADH1B ADH1C	Alcohol dehydrogenase	Cyt	Y Hi Hi				Alcohol (class I), hydroxysteroids, lipid peroxidases
127	ADH4	" "	Cyt					" "
155	ADRB3	Andrenergic recept					Y	
635	BHMT	Methyltransferase	Cyt					Betaine→dimethylglycine, homocyst.→methionine.
1543	CYP1A1	Cytochrome p450	ER	Y				Induced by PAH (cigarette smoke)
1548 **1549** **1553**	CYP2A6 CYP2A7 CYP2A13	Cytochrome p450	ER	Y				Nicotine, coumarin, nitrosamines Unknown Nitrosamine (tobacco). Endo. unknown.
1555 **1556**	CYP2B6 CYP2B7	Cytochrome p450	ER	Y --				Some drugs (e.g. anticancer) Pseudogene.
1558	CYP2C8	Cytochrome p450	ER					Drugs (e.g. taxol-cancer)
1576 **1551**	CYP3A4 CYP3A7	Cytochrome p450	ER	Y				Half of all drugs. Testosterone, estriol formation
1581 **1592** **1595**	CYP7A1 CYP26A1 CYP51A1	Cytochrome p450	ER	Y Hi Hi			Y -- Y	Cholesterol degrad. (rate limiting step) Retinoic acid - regulation gene express. Cholesterol synthesis from lanosterol
2938 **2939** **221357**	GSTA1 GSTA2 GSTA5	Glutathione S-transferase	Cyt	Hi Hi N		Y		Alpha class: Bilirubin, ROS and detoxification. Expression not detected (other studies).
2952	GSTT1	Glutathion s-transf.	Cyt	Low		Y		Theta class: Role in carcinogenesis.
5105	PCK1	Carboxykinase	Cyt		?		Y	Gluconeogen. Insulin reg. Mit isozyme
5475	PPEF1	Phosphatase, Ca						2+ ER-hand Ca binding at C-term.
6519	SLC3A1	a.a. transporter		Low				Cysteine, neutral & dibasic a.a. -kidney
6716	SRD5A2	Steroid reductase						Testosterone, progesterone, corticoster.
6718	AKR1D1	Aldoketo reductase		Y				Bile acids, steroid hormones. Lo→hepatic dysfunct.
6783 **6822**	SULT1E1 SULT2A1	Sulfotransferase	Cyt	Y Hi				Esterone. Control of estrogen receptors. Steroids & bile acids; drugs.
6898	TAT	T NH_3 transferase	Mit	Y	Y			L-tyrosine aminotransferase
7364	UGT2B7	UDP glucuronosyl-transferase 2	ER					Drugs, estrogens. (Detoxification.)
8424	BBOX1	Dioxygenase	Cyt	Mod	Y		?	Carnitine biosynthesis → Transport fatty acid mit.
10891	PPARGC1A	Peroxisome recep.		Y	Y		Y	TF→ energy metab.; cholest. homeo.
29947	DNMT3L	Methyltransferase	Nu					DNA epigenetic transcript. repression.
134147	CMBL	Carboxymethylene-butenolidase hom						Pseudomodas.
151531	UPP2	U phosphorylase 2						Pyrimidine metab. Uridine metab. Mostly kidney.
345275	HSD17B13	Dehydrogenase	Sec					

Table 6: Genes Down Regulated with Shorter Telomere Length - Ten Sample Set.

Entrez Gene ID	Gene Symbol	Gene Type	Histone	Nucleic acid binding	Chromatin packaging and remodeling	Nucloside, nucleotide & nucleic acid metabolism	Cellular Location	Function TF=transcription factor
2074	ERCC6	DNA excision repair	N	Y	N	Y	Nucleus	Transcription-coupled.
4869	NPM1	Nucleolar phospho.	N	N	N	Y	Nucleolus	
5937	RBMS1	RNA binding motif	N	Y	N	Y	Nucleus	DNA repl? Binds ssDNA upstream of c-myc
6165	RPL35A	Ribosomal protein	N	Y	N	N	Cytoplasm	Part of rib. 60S subunit
7182	NR2C2	Nuclear receptor;	N	Y	N	Y		TF. Devel. orphan recept.
8359 8366 8364 8360 8367 8361 8365 8294 8363 8362 8368 8370 554313 121504	HIST1H4A HIST1H4B HIST1H4C HIST1H4D HIST1H4E HIST1H4F HIST1H4H HIST1H4I HIST1H4J HIST1H4K HIST1H4L HIST2H4A HIST2H4B HIST4H4	Histone H4	Y	Y	Y	Y	Nucleus	14 variants of Histone H4 One of four core proteins making up the nucleosome, the core unit of DNA.
8666	EIF3S4	Eukaryotic trans-lation initiation factor	N	Y	N	N		Promotes binding of mRNA and tRNA
9678	PHF14	PHD finger protein	N	Y	N	Y		
9705	ST18	Myelin transcription factor 1	N	N	N	Y		Represses basal transcript act. from target promoters.
9991	ROD1	Regulator of Differentiation	N	Y	N	Y		Suppresses onset of differ-entiation. Nrd1 homolog.
10629	TAF6L	TAF6-like RNA polymerase II	N	N	Y	Y	Nucleus	Part of PCAF complex; acetylates histones.
10919	EHMT2	Methyltransferase	N	Y	Y	Y	Nucleus	Methylates histone H3, H1
11317	RBPSUHL	Recombining binding protein	N	Y	N	Y		Transcr. factor w/EBNA2? Suppressor of hairless.
23521	RPL13A	Ribosomal protein	N	Y	N	N	Cyt	Part of 60S subunit.
29396	unassigned	Transformer-2 alpha	N	Y	N	Y		Myelin transcription factor
54993	ZSCAN2	Zinc finger and SCAN domain	N	Y	N	Y	Nucleus	Transcript regulation post-meiotic spermatogenesis?
55198	unassigned	DIP13 Beta	N	Y	N	N	Mem, nuc	Regulation of cell prolif.
56907	SPIRE1	Spire homolog	N	N	N	Y		Actin organization.
64843	ISL2	ISL2 transcript factor	N	Y	N	Y	Nucleus	TF; subclass of moto-neurons of spinal cord.
80345	ZNF435	Zinc finger protein	N	Y	N	Y		Transcription regulation?
126820	WDR63	Testis develop. prot.	N	N	N	Y		
171017	ZNF384	Zinc finger protein	N	Y	N	Y	Nucl.	TF. MMP. Discontinued?

Appendix C Excel Codes

Sample Set A - Ten Liver Sample Dataset

Preprocessing: Adjustments, modifications and averaging of data, and ANCOVA macro.

(Note: no similar processing was needed for Sample Set B - GEO GSE9588.)

Data Processing Overview:

1. Correct Gene ID codes.
2. Remove data points which fail the quality control.
3. Raise all intensities to the minimum, the Negative TrimMean Control.
4. Transform data to log 2, and then average. If a data point is missing, use the other value. If both are missing, mark "Missing." If the replicas differ by more than 8 fold, remove the average data point.
5. Determine the number of data points not omitted. (If more than one missing, omit gene.)
6. Determine the skew of the data for a single gene. Remove that gene if the absolute value of the skewness is greater than two times the Standard Error of Skewness (SES=$\sqrt{6/n}$)
7. Determine the kurtosis of the data for a single gene. Remove that gene if the absolute value of the kurtosis is greater than two times the Standard Error of Kurtosis (SEK = $\sqrt{24/n}$).
8. Copy data to a new sheet.
9. Check data.
10. Run ANCOVA macro for each gene using XL-STAT. (www.xlstat.com)
11. Check that probe IDs match.
12. Search DAVID for equivalent reference sequence gene identification numbers.
13. Find ANCOVA genes in a specific list of inflammation and ROS genes, for closer analysis.
14. Extract the slope of the ANCOVA line, to determine the direction of expression.

1. Remove .1 suffix from the ACCN#, for recognition by DAVID, etc.
   =LEFT(B2,LEN(B2)-2)

2. Mark microarray values with quality control "C", "M" or "I" as "Missing"

3. Raise numbers to the Negative TrimMean Control for that specific sample.
   =IF('LGSMI All'!H2="Missing","Missing",IF('LGSMI All'!H2<0.679853734,0.679853734,'LGSMI All'!H2))

4. Take the log base 2 of the numbers, and then average them.
   If both values are missing, mark "Missing"
   If one of the values is missing, the "average" is log2 of the other value.
   If the variance between the two values is greater than 8 fold, mark "Variance".
   For all others, take the log base 2, and then average the two numbers after.
   Result: 7,257 had a variance greater than 8x
   Result: 2,200 had both data points missing.

   =IF(AND('To TrimMean'!F2="Missing",'To TrimMean'!G2="Missing"),"Missing", IF('To TrimMean'!F2="Missing",LOG('To TrimMean'!G2,2), IF('To TrimMean'!G2="Missing",LOG('To TrimMean'!F2,2),IF(ABS(LOG('To TrimMean'!F2,2)-LOG('To TrimMean'!G2,2))>3,"Variance",AVERAGE(LOG('To TrimMean'!F2,2),LOG('To TrimMean'!G2,2))))))

5. Determine the number of data points not omitted.
   = Count(F2:O2)

6. Determine the skewness of the distribution. Mark as 99 if the number of valid data points is less than 4.
   =IF(P3<4,99,SKEW(F3:O3))

7. Determine the kurtosis of the distribution. Mark as 99 if the number of valid data points is less than 5.
   =IF(P2<5,99,KURT(F2:O2))

8. Copy the values and paste into a new sheet. Convert the "Missing" and "Variance" to "NA" for the ANCOVA analysis using the Replace tool, which is what XL-STAT requires for missing data points.
   Paste special → values

9. Check the data that all values are within -2 and 10 (which is the log2 of 0.25 and 1064) or are marked NA (no value - "Missing" or "Variance").
   =OR(AND('Y Variable'!G2<10,'Y Variable'!G2>-2),'Y Variable'!G2="NA")
   Results: One value was 10.005. All others confirmed. Confirmation data omitted.

10. XL-STAT Macro for ANCOVA (Age-Disease and Telomere-Disease)
Adjusted slightly for each run. i = 1 to 49584. Run in sets of 3,000 to prevent RAM-related computer crashes.

```
Sub ANCOVADisease()
   Dim i As Long

   For i = 49001 To 49584

      Call LoadRunANCO(Sheets("Y Variable").Range("F" & 1 + i & ":O" & 1 + i), _
                  Sheets("X Variable").Range("F2:O2"), _
                  Sheets("X Variable").Range("F3:O3"), _
                  NoScreenUpdating:=True)

      DoEvents
      With Sheets("Results").[E1]
         .Offset(i) = ActiveSheet.Range("C37")
         .Offset(i, 1) = ActiveSheet.Range("C39")
         .Offset(i, 2) = ActiveSheet.Range("C40")
         .Offset(i, 3) = ActiveSheet.Range("C41")
         .Offset(i, 4) = ActiveSheet.Range("C42")
         .Offset(i, 5) = ActiveSheet.Range("C43")
         .Offset(i, 6) = ActiveSheet.Range("C44")
         .Offset(i, 7) = ActiveSheet.Range("E55")
         .Offset(i, 8) = ActiveSheet.Range("E56")
         .Offset(i, 9) = ActiveSheet.Range("F55")
         .Offset(i, 10) = ActiveSheet.Range("G55")
         .Offset(i, 12) = ActiveSheet.Range("F64")
         .Offset(i, 13) = ActiveSheet.Range("G64")
         .Offset(i, 14) = ActiveSheet.Range("F65")
         .Offset(i, 15) = ActiveSheet.Range("G65")
         .Offset(i, 16) = ActiveSheet.Range("F66")
         .Offset(i, 17) = ActiveSheet.Range("G66")
         .Offset(i, 19) = ActiveSheet.Range("F72")
         .Offset(i, 20) = ActiveSheet.Range("G72")
         .Offset(i, 21) = ActiveSheet.Range("F73")
         .Offset(i, 22) = ActiveSheet.Range("G73")
         .Offset(i, 23) = ActiveSheet.Range("F74")
         .Offset(i, 24) = ActiveSheet.Range("G74")
         .Offset(i, 25) = ActiveSheet.Range("B92")

         Application.DisplayAlerts = False
         ActiveSheet.Delete

      End With

   Next i

   Application.DisplayAlerts = True

End Sub
```

11. Check that the probe analyzed in the results (given in the equation of the model) matches the probe and description listed, and matches the same probe on the line in the y variable. The probe names all start with GE and have at least 7 characters. Filter to examine any that are marked "False." Result: only those not analyzed (i.e. n<8) came up "False."
    =LEFT(AD2,7)=LEFT(D2,7)
    =LEFT(AD2,7)=LEFT('Y Variable'!F2,7)

12. Search DAVID for equivalent mRNA reference sequences in non NM_ gene ascension numbers. Get the NM_ number and the gene description. Mark those descriptions which are unusable: ORF, Hypothetical proteins, Hypothetical genes, KIAA, Incyte unique probes and LOC. Copy and paste these into the ACCN# column, and sort by this column. Then use the following code to copy and paste the new ID and description into that column.
    =IF(AND(B2=B4,NOT(B4="")),E4,IF(B2=B3,E3,""))

13. Find the ANCOVA results from a list specific genes related to inflammation and reactive oxidative species (ROS).
    a. Copy and paste the ANCOVA results into a new file, to avoid inadvertent contamination of the results.
    b. Insert columns for the acronym and group number.
    c. Set up the blank cells to copy the line below it if the genes are the same.
       =IF(I2=I3,E3,"")
    d. Cut and paste the list of genes of interest, so that the GE… ID numbers are in the same column.
    e. Sort the data by gene (GE…)
    f. Set up a column to identify if the genes are the same.
       =I2=I3
    g. Auto filter by that column, those marked "True". Copy and paste those genes into a new worksheet. State which ANCOVA analysis they come from (Age-Tel, Age-Dis or Tel-Dis).

14. Extract the slopes from the overall equation of the model in the Age-Tel group. (There are no spaces in the original equation. They were added here for readability. LEN (length) = # of characters in a cell. LEFT & RIGHT = Take the first x number of characters starting from the left or right side, respectively. MID = Take the middle characters. FIND = Find the first character (or word); returns the number of characters.
    a. AD2-Start: GE545180 = -12.0993263871198 + 0.302036367147246*Age + 1.40056410472251*Telomere Length - 0.034985998351747*Age*Telomere Length
    b. AJ2-Right side: =RIGHT(AD2,LEN(AD2)-FIND("=",AD2,3))
    c. AK2-Left 2 terms: =LEFT(AJ2,FIND("*",AJ2,1)-1)
    d. AL2-Left term: =LEFT(AK2,FIND(".",AK2,6)-3)
    e. AM2-Account for single vs. double digit 2nd term (10.00 vs. 9.00): =RIGHT(AL2,1)
    f. AO2-Intercept slope: =IF(OR(AM2="+",AM2="-"),LEFT(AL2,LEN(AL2)-1), AL2)
    g. AP2-2nd term: =RIGHT(AK2,LEN(AK2)-LEN(AO2))
    h. AQ2-Check sign of the term, to remove the '+'.
    i. AS2-Age slope: =IF(AQ2="-",AP2,RIGHT(AP2,LEN(AP2)-1))
    j. AT2-Last two terms: =RIGHT(AJ2,LEN(AJ2)-LEN(AK2)-4)
    k. AU2-Sign of third term, to remove the '+': =LEFT(AT2,1)
    l. AW2-Telomere Length: =IF(AU2="-", LEFT(AT2,FIND("*",AT2,2)-1), MID(AT2,2,FIND("*",AT2,2)-2))
    m. AX2-Last term: =RIGHT(AT2,LEN(AT2)-LEN(AW2)-17)
    n. AY2-Sign of last term, to remove the '+': =LEFT(AX2,1)
    o. BA-Age-Telomere: =IF(AY2="-", LEFT(AX2,FIND("*",AX2,2)-1), MID(AX2,2,FIND("*",AX2,2)-1))
    p. Convert from formulae to numbers, to stabilize:
        i. Copy → Edit → Paste Special → Value
        ii. Convert from text to a number value ('!' icon next to cell)