National Center for Forensic Science
University of Central Florida
P.O. Box 162367 · Orlando, FL 32826
Phone: 407.823.4041
Fax: 407.823.4042
Web site: http://www.ncfs.org/

Biological Evidence

The Biological Age of a Bloodstain Donor

FINAL REPORT

May 27, 2014

Department of Justice, National Institute of justice
Award Number: 2009-DN-BX-K179
(1 October 2009 – 31 May 2014)

Principal Investigator:
Jack Ballantyne, PhD
Professor
Department of Chemistry
Associate Director for Research
National Center for Forensic Science
P.O. Box 162367
Orlando, FL 32816-2366
Phone: (407) 823 4440
Fax: (407) 823 4042
e-mail: Jack.Ballantyne@ucf.edu

ABSTRACT

Biological samples from crime scenes are routinely analyzed in an attempt to determine the identity of the depositor. Forensic DNA analysis techniques employ a battery of non-coding STR loci to determine individuality by direct comparison of a crime scene sample and a reference sample. No information about the physical characteristics of the donor can be gleaned from the STR profiles *per se*. However the ability to predict some key physical features of an individual by assaying appropriate specific biomarkers from crime scene samples would greatly aid criminal investigations. Recent progress has been made in identifying biomarkers associated with physical characteristics such as hair-, skin- and eye- pigmentation and bio-geographic ancestry but other physical features may be similarly amenable to genetic analysis.

In the current work we sought to identify age-specific biomarkers with the ultimate aim of designing prototype assays that may be used to predict the biological age of a bloodstain sample donor. An individual's age is innately associated with developmental genetic changes. Several metabolic and biochemical processes within the body are key to specific developmental stages in a person's life and these are regulated by specific genes. Thus the possibility exists of finding an association between patterns of mRNA expression of such regulatory genes (and the genes that they regulate) and specific stages of life. We took a two-pronged approach to biomarker identification. The first strategy was to analyze the entire transcriptome of selected samples of different ages, using deep sequencing RNA-Seq technology, in order to identify candidate genes that show different expression with respect to the age of the donor. The second approach involved targeting specific candidate genes that are, or are likely to be, involved in regulation of age-related processes based upon *a priori* understanding of the physiology and biochemistry of human development.

In this study, we evaluated over 500 mRNA and 1,100 miRNA candidates for the identification of potential age biomarkers for the determination of the biological age (newborn, toddler, child, adolescent, adult, mature adult and elderly) of a bloodstain donor. We have successfully identified a set of 11 novel mRNA gene candidates, that along with two housekeeping genes (B2M, S15) and our previously identified newborn markers (HBG1n and HBG2n), permit an identification of an individual as belonging to one of 5 major age group classifications: (i) newborn, (ii) infant-toddler, (iii) child-adolescent, (iv) adult/mature adult and (v) elderly. The analysis of these ~15 mRNA gene candidates can be performed using four qPCR assays (with three of the four assays at this time only requiring the use of unlabeled PCR primers and SYBR green reagents). Additionally, we have employed the use of logistic regression analysis to provide statically sound 'predictions' of the biological ages in three of the four assays. Proposed workflows for the classification of unknown samples have been developed. In addition to the mRNA profiling assays developed in this study, we identified (and continue to evaluate) 36 miRNA biomarkers as potential age biomarkers. We are hopeful that with continued analysis, we can develop suitable miRNA profiling assays for age determination that may serve to further refine or support age classifications made by the mRNA profiling assays.

While all assays developed in the current study require additional validation before they could be used routinely, initial results demonstrate the potential use of RNA profiling for the determination of the biological age of bloodstain donors.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

1. Crime scene DNA STR profiles provide no information about any innate physical characteristics of the depositor. In many criminal and unidentified remains cases the investigation would be greatly aided if it was possible to predict the physical appearance of a sample donor, particularly in the absence of a witness or other relevant information about the donor.

2. Eyewitness accounts include descriptions of features that will be, to varying degrees, dependent upon inherent underlying molecular genetic phenomenon, as opposed to environmental influences such as scars, tattoos, trauma etc. Inherent physical features commonly described in eyewitness reports include an individual's gender, ethnicity, skin/hair/eye color, height, build, facial features and age. In the current work, we sought to determine whether determination of a donor's biological age could be a physical characteristic amenable to molecular genetic dissection and used by the forensic community.

3. The human body ages over time: initial growth and development takes place during early childhood, which is then followed in sequential order by maturation into an adult during puberty, the adulthood years (which can significantly determine the extent of degenerative changes that will occur during old age) and finally middle age, old age and death. These obvious changes are governed by a network of not so obvious, highly complex metabolic and genetic systems. These systems are activated or repressed during a person's lifetime and thus the metabolic state of these systems is expected to be representative of particular developmental ages. Therefore, it would seem reasonable to postulate that some of these changes in gene expression over the life cycle could be identified and used for a determination of the biological age of a bloodstain donor.

4. Several unsuccessful attempts have been made to try to correlate some of the stochastic changes (cellular senescence, the general wear and tear of cells, protein, somatic mutations, oxidative DNA damage, mitochondrial damage and telomere length shortening) with the biological age of the donor of forensic specimens. For example, telomere shortening was found (by us and others) not to correlate with chronological age in a variety of human tissues. Consequently, in this study, we concentrated on identifying biomarker correlates of the systemic processes involved during developmentally regulated changes in gene expression and metabolism during the human aging process from birth through the late adult stage (i.e. "elderly").

5. Previous research in our laboratory resulted in the discovery of markers for distinguishing neonatal samples from other age groups using mRNA expression profiling of novel hemoglobin G (HBG) mRNA isoforms. We had also obtained preliminary evidence in non-neonates by mRNA screening of blood samples from individuals of different age groups that the relative expression of COL1A2, IGFBP3 and S15 might provide low resolution differentiation of individuals into separate age groups in dried bloodstains (unpublished observations). Therefore

we had preliminary indications that mRNA profiling for age determination beyond the newborn stage would be a feasible task.

6. The principal aim of this project was to perform an exhaustive and focused search and testing of the expression of genes in order to identify differentially expressed mRNA transcripts that might be used to predict a bloodstain donor's biological age. For the purposes of this study, age categories were defined as follows: newborn/infant (birth - < 2 years); toddler (2-6 years); child (7-12 years); adolescent (13 – 17 years); adult (18 – 45 years); mature adult (46 – 65 years); and elderly (≥ 65 years). These categories generally correlate with known psychosocial and biological changes known to occur throughout the human development life cycle.

7. We took a two-pronged approach to mRNA biomarker identification. The first strategy included the analysis of entire transcriptomes of selected samples of different ages, using deep sequencing mRNA-Seq technology, in order to identify candidate genes that exhibit different expression levels with respect to the age of the donor. The second approach involved targeting specific candidate genes that are, or are likely to be, involved in regulation of age-related processes based upon *a priori* understanding of the physiology and biochemistry of human development. Both approaches were ultimately successful in the identification of potential age biomarkers. Over 500 mRNA gene candidates (~20% identified through RNA-seq) were evaluated in the current study.

8. We successfully developed three mRNA profiling real time PCR assays (qPCR) for the identification of "newborn/toddler", "younger/older" and "elderly" assays: 1) newborn/toddler (IGFBP3, COL1A2), 2) younger/older (CCR7, RBBP8-v3, GFI1, GSTM3, SENP1) and 3) elderly (IGJ, UQCRFS1, SENP1, FCGR2B, NDUFB11). Novel logistic regression (logR) models were developed to aid in the interpretation of the expression data for three of these assays. The logR models permit the incorporation of multi-gene candidates into a single model thereby increasing the predictive power of the assay compared to using individual markers. These models provide powerful age "prediction' or classifications. The use of these three qPCR assays, in addition to our previously developed newborn qPCR assay (HBG1n, HBG2n), provide the ability to classify a bloodstain into one of 5 age categories: 1) newborn, 2) infant/toddler, 3) child/adolescent, 4) adult/mature adult and 5) elderly (note: the placement of adolescent samples has not be definitively identified at the time of this report and could result in an alternate binary classification of "child" and "adolescent/adult/mature adult" rather than "child/adolescent" and "adult/mature adult").

9. The younger/older and elderly assays are performed using a simple SYBR green-based qPCR reaction, thereby eliminating the need for costly fluorescently labeled probes. Additionally, we have developed a simple and inexpensive pre-amplification method to improve the sensitivity of assays involving lower abundance candidates. This pre-amplification step is utilized in the newborn/toddler (IGFBP3/COL1A2) assay.

10. In addition to the >500 mRNA gene candidates, we also tested > 1,100 miRNA biomarkers using individual primer assay screening as well as the use of miRNA array plates. From this analysis, we have identified 36 potential age-correlated miRNA candidates that warrant further

investigation. The evaluation of these miRNA candidates is ongoing at the present time. We will develop logistic regression analysis models with the best subset of miRNA candidates in the hope that they will improve, either alone or in combination with the mRNA assays, the predictive power of our age classification scheme.

11. While all of the reported biomarker assays require more extensive validation, the results of the current study support the concept of using RNA profiling assays as a means to accurately categorize the biological age of a bloodstain donor into one of five major age groups.

I. INTRODUCTION

A. Statement of the Problem

Crime scene DNA STR profiles provide no information about any innate physical characteristics of the depositor. In many criminal and unidentified remains cases the investigation would be greatly aided if it was possible to predict the physical appearance of a sample donor, particularly in the absence of a witness or other relevant information about the donor. One such biometric is age, which is one of the primary characteristics that we notice in a person and is governed by processes that are largely developmental in nature and, importantly should be amenable to molecular genetic analysis. Therefore, in the current work we sought to identify age-specific biomarkers with the ultimate aim of designing assays that may be used to predict the biological age of a sample donor.

B. Literature Review

The ability to predict some key physical features of an individual by assaying appropriate specific biomarkers from crime scene samples would greatly aid criminal investigations. Ideally this genetic eyewitness would at least recapitulate, if not improve, the identifying characteristics commonly described in eyewitness reports. Eyewitness accounts include descriptions of features that will be, to varying degrees, dependent upon inherent underlying molecular genetic phenomenon, as opposed to environmental influences such as scars, tattoos, trauma etc. Inherent physical features commonly described in eyewitness reports include an individual's gender, ethnicity, skin/hair/eye color, height, build, facial features and age. Gender is already a routine part of DNA analysis due to the incorporation of the AMEL locus into commercial DNA testing kits [1-3] and bio-geographic testing for ethnicity prediction is now available [4-7]. Significant progress has been made in identifying pigmentation biomarkers for eye [8-22], skin [14,16,17,23-31] and hair [17,24,30-34] color, and forensic assays incorporating some of these biomarkers are currently under development. Of the remaining features height, although highly heritable (~80%), represents a significant challenge in that it is likely that, based upon genome wide association studies, a very large number of genes of small effect (>100) or a small number of rare (i.e. 'private' to the individual or his relatives) genes of large effect will be required to explain the heritability component of the variance in height [35-43]. Thus further progress in our fundamental understanding of height regulation in humans is required before height becomes amenable to forensic analysis. Build or body mass index (BMI) is related to height (BMI = weight (kg)/height $(m)^2$) but is further complicated by the large environmental contribution to the weight component [44-48]. It is questionable whether BMI will ever be useful for molecular phenotyping in the forensic context. Facial features are determinative of an individual's appearance and, although apparently exceedingly complex at the genetic level, are under strong genetic control as evidenced by the nearly identical facial appearance of monozygotic twins. The developmental genetics of craniofacial development is currently the object of intense investigation [49-53]. The last of the aforementioned features involves the chronological age of an individual. For reasons we describe below, we believe this phenotype to be the next physical characteristic that is likely to be amenable to molecular genetic dissection and use by the forensic community.

Theories and Underlying Bio-molecular Mechanisms of Ageing

The human body ages over time: initial growth and development takes place during early childhood, which is then followed in sequential order by maturation into an adult during puberty, the adulthood years (which can significantly determine the extent of degenerative changes that will occur during old age) and finally middle age, old age and death [54]. These obvious changes are governed by a network of not so obvious, highly complex metabolic and genetic systems. These systems are activated or repressed during a person's lifetime and thus the metabolic state of these systems is expected to be representative of particular developmental ages. Aging in humans is often studied within the context of two separate disciplines, developmental biology and gerontology [54-56]. Developmental biology deals with the changes that occur during the early formative year of life from fertilization until maturation during puberty while gerontology deals mainly with the degenerative changes that occur during post-pubertal adulthood and old age.

Current belief is that aging is a process that gradually leads to the decline in the fitness of an organism and involves several interacting physiological and biochemical networks [55-57]. Common theories of aging have led to the concept that both systemic and stochastic changes occur in the body over time [54,55]. Systemic changes pertain to the developmentally regulated changes in gene expression and metabolism that have evolved in humans. Stochastic processes on the other hand are un-programmed and occur randomly and continuously over an individual's life time. These latter processes are mainly manifested as an increase in damaged, altered or degraded molecular products leading to cellular malfunction. Several mechanisms have been proposed to be the cause of aging. These include, *inter alia*, cellular senescence, the general wear and tear of cells [58], protein modifications such as glycation [59,60] and oxidation [61], somatic mutations [62,63], oxidative DNA damage [64-66], mitochondrial damage [67] and telomere length shortening [68]. It is not clear at this stage whether any or all of these are the cause of aging or the result of other as-yet-unknown age-causing processes [54,55,57].

Forensically Relevant Developments in Ageing Research: Biological and Technological

Several unsuccessful attempts have been made to try to correlate some of the aforementioned stochastic changes with the biological age of the donor of forensic specimens. For example, telomere shortening was found not to correlate with chronological age in a variety of human tissue [69,70]. Consequently, in this study, we concentrated on studying the systemic processes involved in aging. These include the myriad of biochemical and molecular networks that function in synchronicity with each other and lead to human development and aging. Research has shown that endocrine pathways associated with the hypothalamic-pituitary-gonadal axis plays a key role during puberty [71-73]. Genes involved in the regulation of these pathways would be excellent candidates for differentiating between childhood and adulthood. Leptin and its receptor have also been implicated in the regulation of puberty and may also be viable candidates [73]. MITF, BCL2 and KitL/c-Kit [74-76] are key molecular modulators of the process of cellular differentiation and play a significant role in the development of hematopoietic cells and osteoclasts. Delineating the expression of such genes as a function of age could yield worthwhile pre- or post-pubertal candidates. In older aged individuals, genes involved in the repair of DNA are expected to be up-regulated in response to the heightened amount of endogenous and environmentally induced DNA damage experienced by such individuals [77].

Recent years have seen explosive advances in whole genome analysis techniques. 'Next Generation Sequencing' (NGS) technologies have revolutionized the field of bioinformatics and

have completely changed our understanding of genome biology [78-83]. These deep-sequencing techniques permit the researcher to sample simultaneously individual DNA sequences from heterogeneous mixed samples, thus providing quantitative and qualitative information about the genome under specific experimental or environmental conditions. RNA-Seq refers to the process of deep sequencing of RNA transcripts using NGS techniques that, unlike standard micro-array analysis, permits the mapping and quantification of whole transcriptomes, including the presence and abundance of mRNA isoforms [79].

Whole genome studies including the use of NGS methods have aided our understanding of the connection between age and genes. McCarrol et al. (2004) [84] studied gene expression across four species and found that most age regulated gene expression was species specific. Zhan et al. (2006) [85] compared the expression profiles in three human tissues (kidney, brain and muscle) and found six genetic pathways that demonstrated differential expression with age. They found an increase in the expression of several genes encoding components of, or regulators of, the extracellular matrix, the cellular growth pathway, complement activation and cytosolic ribosomes. Two multi-gene systems (chloride transport and subunits of the mitochondrial electron transport chain) showed a decrease in the expression of the constituent genes with age. Interestingly, changes in the expression of genes of the mitochondrial electron chain pathway have been implicated in a number of aging studies [86-89]. Recently, fifteen different metabolic pathways were shown to exhibit common age regulated changes between humans and mice [86]. A gene regulation network database has been created (AGEMAP) that lists and maps the thousands of genes known to be age-regulated in different tissues of mice [88]. Three different tissue classes could be discerned (neural, vascular and steroid responsive) depending upon the kind of age regulated expression patterns they exhibited. These and other studies indicate that genetic factors play a role in the ageing processes. Blood has not been studied extensively for gene expression changes with age due to the heterogeneous nature of its constituent cell population (but see below). However it is probably the most common biological tissue encountered at crime scenes and could turn out to be an excellent matrix for age biomarker identification since a number of different physiological systems could be monitored simultaneously.

The current study relies on a number of fundamental assumptions. These are (i) that differential gene expression, as exemplified by mRNA transcript levels, will occur throughout human development in human tissues; (ii) that blood, as essentially a self–renewing tissue, will still exhibit age related differences in gene expression due to the ageing of the stem cell compartment responsible for blood cell production; (iii) that RNA is sufficiently stable in dried stains for forensic analysis. Empirical support for these assumptions has been obtained and will be briefly summarized in the remainder of this section.

A number of studies employing micro arrays have demonstrated age related differences in mRNA levels in a variety of different human tissues including brain [90-94], kidney [95,96], muscle [85,97-100], liver [101], eye [102-104], skin [105] and blood [106,107]. With respect to the tissue of interest in this proposal, namely blood, Visala et al. (2003) [107] found 76 genes that displayed more than 2.5 fold differences in expression levels in peripheral blood lymphocytes between old and young individuals. Specifically signal transduction and mitochondrial respiration genes were up-regulated and heat shock response and cell survival genes were down regulated in old compared to young individuals.

All blood cells originate from hematopoietic stem cells (HSC) during a process of regulated development in the bone marrow in which pluripotent HSC become increasingly differentiated to

mature effector cells (e.g. erythrocytes, lymphocytes) through a series of progressively more committed progenitor cells [108]. Advancing age is accompanied by pathophysiological changes in the hematopoietic system that suggests a loss of homeostatic control and the inability of HSC to properly and efficiently reconstitute the blood cell compartments [109-111]. Studies have directly demonstrated decline in HSC function with age. Accumulation of DNA damage appears to be one mechanism that could cause functional decline [112,113]. Moreover, Chambers et al. (2007) [114] demonstrated that old HSCs showed higher order changes in overall gene expression compared with younger HSCs in that genes or chromosomal regions that were silenced in young HSCs were now turned on and, conversely, other genes were expressed at lower levels than in younger HSC cells. Thus RNA expression in blood is expected to differ according to the age of the individual. In addition, since blood bathes all tissues, age related changes in non-self renewing tissues is likely to be represented to varying degrees in the blood due to secretion from, and sloughing off, of cells from these tissues. As a result the mRNA composition of blood is comprised of RNA from hematopoietic cells and from other non-blood tissues. Thus, for the aforementioned reasons, there is good reason to be confident in the ability to detect chronological age mRNA biomarkers in blood.

Previous research in our laboratory resulted in the discovery of markers for distinguishing neonatal samples using mRNA expression profiling of novel hemoglobin G (HBG) mRNA isoforms [115,116]. We have also obtained preliminary evidence in non-neonates by mRNA screening of blood samples from individuals of different age groups that the relative expression of COL1A2, IGFBP3, HBG1/2n and S15 can provide low resolution differentiation of individuals into 4 separate age groups in dried bloodstains (unpublished observations). Kerschan-Schindl et al. [117] reported that the c-terminal telopeptide of type I collagen, was increased in elderly subjects. This led us to an investigation of genes known to be associated with bone development and we identified the collagen, type I, alpha 2 gene (COL1A2) [117] as a potential biomarker of ageing (unpublished observations). Additionally we identified the insulin-like growth factor binding protein 3 gene (IGFBP3) [118,119], a regulator of cell somatic growth and cellular proliferation, as another potential age biomarker (unpublished observations). S15 encodes a structural RNA present on the ribosome and is regarded as a housekeeping gene since it is expressed in all tissues [120]. This latter species is used to normalize the expression of COL1A2 and IGFBP3 in dried bloodstains in a triplex qRT-PCR assay (Figure S1, *requires viewing in color for proper resolution*).

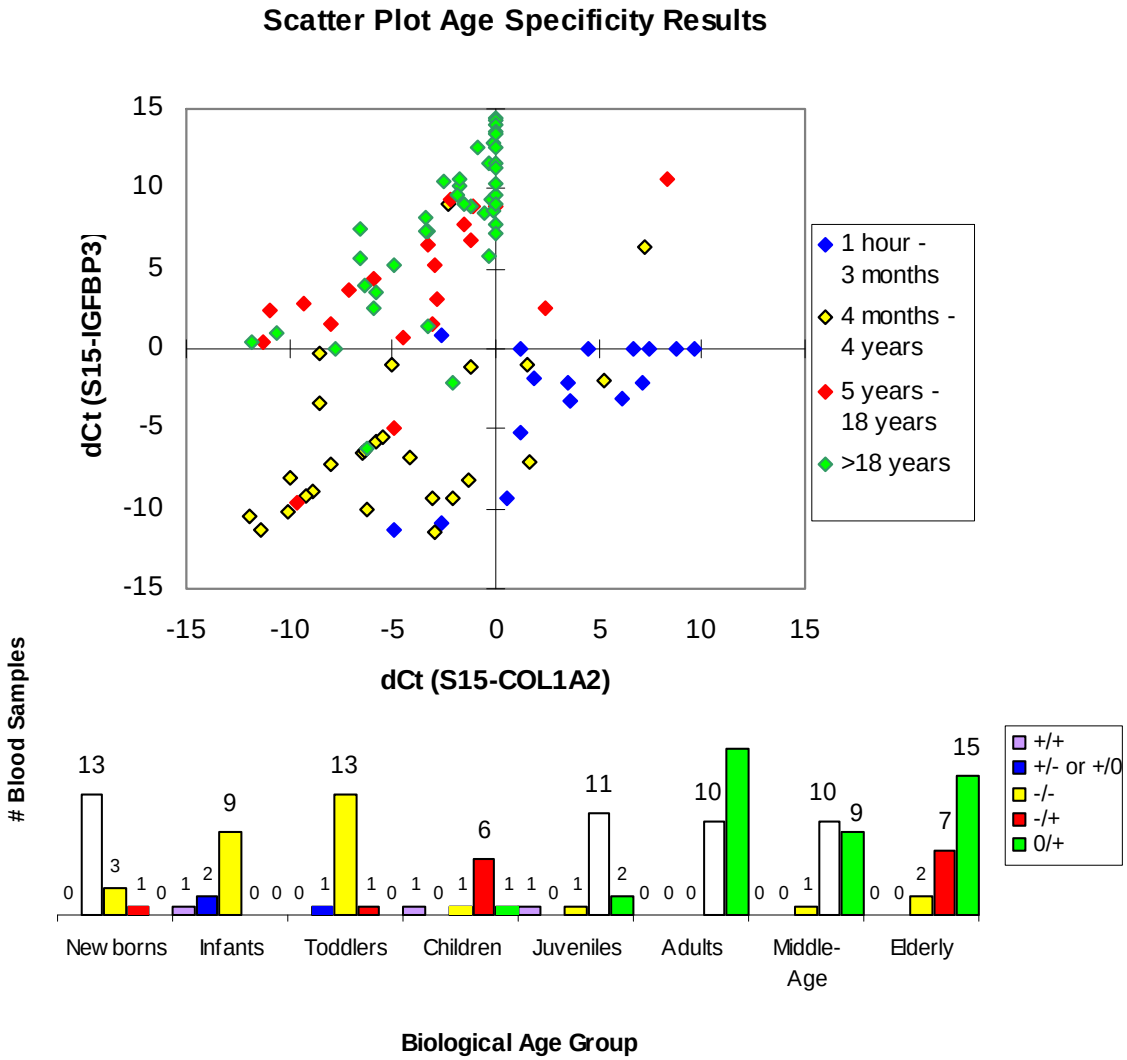


Figure S1. Scatter plots in which each sample's dCt (S15-COL1A2) vs. dCt (S15-IGFBP3) is graphed into one of the four quadrants. As illustrated, individuals can be categorized into four (albeit overlapping) age groups; newborns (1-hour to 3-months), located on the axis to the right of the origin (+/0) and in the lower right quadrant (+/-), infants/toddlers (4-months to 4-years), located in the lower left quadrant (-/-), children/juveniles/adults/middle-age/elderly (5-years to 102-years), located in the upper left quadrant (-/+) and on the axis above the origin (0/+). (Lower Panel) Graphical representation of the ddCt scatter plot results indicating the number of individuals within each biological age and the generated ddCt results.

The current study was focused on furthering this research and discovering markers for all major stages of life and developing a biomarker panel that can be used by the forensic community to predict the biological/chronological age of the bloodstain depositor. The current study differed from our previous work in that a much more exhaustive and focused search and testing of genes from the known biochemical pathways was carried out. In addition, the mRNA-Seq technique of gene discovery offers a number of significant advantages over the previously

used hybridization technology that was expected to facilitate the discovery of age specific genes or RNA isoforms.

Finally, we and others have demonstrated that mRNA and miRNA molecules are stable in dried bloodstains, including those exposed to the range of environmental insults experienced by forensic stains [121-130]. Generally, RNA is generally no more or less susceptible to environmentally induced profile loss than DNA, probably in part due to the fact that each cell contains thousands of RNA transcripts from a particular locus compared to two for autosomal DNA markers.

C. Statement of Hypothesis or Rationale for the Research

The principal aim of this project was to identify genes that show age related changes in their RNA expression levels. Our own previous studies resulted in the identification of mRNA markers specific for neonates. However, broader spectrum markers that differentiate samples over a wider range of ages, from early childhood through old age, have not yet been characterized. We targeted specific age regulatory genetic pathways (candidate gene approach) as well as investigated the human transcriptome of blood cells as a function of biological age using a 'Next Generation Sequencing' platform (gene discovery approach). Appropriate assays and biostatical tools would be developed for any specific biomarkers identified.

II. METHODS

Preparation of Body Fluid Stains

Bloodstains were collected from individuals of varying (but known) age and gender. The samples range in age from neonates to the elderly (i.e. newborn/1 day - 100 years of age) and were categorized into different groups according to developmental age: newborn/infant (< 2 years of age), toddler (2-6 years), child (7-12 years), teenager/adolescent (13-17 years), adults (18-45 years), middle-aged (45–60 years) and elderly (>60 years). Blood samples were collected by venipuncture into vacutainers and 50 µl aliquots were placed onto cotton cloth and dried at room temperature. All samples were stored at -20^{o}C until needed. A 50µl stain or a single cotton swab was used for RNA isolation unless otherwise stated.

Some of the obtained blood samples were collected from donors by Florida Hospital (Orlando, FL) after receiving exemption from the Hospital's Review Board. We also received additional blood samples from Tampa General and Arnold Palmer Hospital for Children (Orlando Health, Orlando FL). All samples were collected in accordance with the procedures approved by the Hospitals and the Institutional Review Board at the University of Central Florida. Additional blood samples (ages 19 - 70 years old; male and female; 1 vacutainer (~10 mL)) were purchased from Bioreclamation Inc. (New York).

RNA Isolation

Total RNA was extracted from blood, semen, saliva, vaginal secretions and menstrual blood and skin with guanidine isothiocyanate-phenol:chloroform and precipitated with isopropanol [131]. Briefly, 500 µl of pre-heated (56^{o}C for 10 minutes) denaturing solution (4M guanidine isothiocyanate, 0.02M sodium citrate, 0.5% sarkosyl, 0.1M β-mercaptoethanol) was added to a 1.5mL Safe Lock tube extraction tube (Eppendorf, Westbury, NY) containing the stain or swab. The samples were incubated at 56^{o}C for 30 minutes. The swab or stain pieces were then placed into a DNA IQ^{TM} spin basket (Promega, Madison, WI), re-inserted back into the original extraction tube, and centrifuged at 14,000 rpm (16,000 x g) for 5 minutes. After centrifugation, the basket with swab/stain pieces was discarded. To each extract the following was added: 50 µl 2 M sodium acetate and 600µl acid phenol:chloroform (5:1), pH 4.5 (Ambion/Life Technologies, Grand Island, NY). The samples were placed at 4^{o}C for 30 minutes to separate the layers and then centrifuged for 20 minutes at 14,000 rpm (16,000 x g). The RNA-containing top aqueous layer was transferred to a new 1.5ml microcentrifuge tube, to which 2µl of GlycoBlue^{TM} glycogen carrier (Applied Biosystems/Ambion) and 500 µl of isopropanol were added. RNA was precipitated for 1 hour at -20^{o}C. The extracts were then centrifuged at 14,000 rpm (16,000 x g). The supernatant was removed and the pellet was washed with 900 µl of 75% ethanol/25% DEPC-treated water. Following a centrifugation for 10 minutes at 14,000 rpm (16,000 x g), the supernatant was removed and the pellet dried using vacuum centrifugation (56^{o}C) for 3 minutes. Twenty microliters of pre-heated (60^{o}C for 5 minutes) nuclease free water (Ambion/Life Technologies) was added to each sample followed by an incubation at 60^{o}C for 10

minutes. Samples were used immediately or stored at -20°C until needed. An extraction blank (reagents only, no sample) was included with each extraction as a negative control.

DNase I digestion

DNase digestion was performed using the Turbo DNA-*free*™ kit (Ambion/Life Technologies) according to the manufacturer's protocol. Briefly, 2U units of TURBO™ DNase (2U/μl) and 2 μl (0.1 volume) of Turbo DNase I Buffer (10X) were added to each RNA extract (20 μl) and incubated at 37°C for 30 minutes. The DNase was inactivated by the addition of 2.3 μl (0.1 volume) of the supplied DNase Inactivation Reagent. The samples were incubated at room temperature (occasional mixing) for 5 minutes. The samples were centrifuged at 10,000 x *g* for 1.5 minutes and then the supernatant was transferred to a new 1.5mL tube. The samples were used immediately or stored at -20°C until needed.

RNA quantitation

RNA extracts were quantitated with Quant-iT™ RiboGreen® RNA Kit (Invitrogen, Carlsbad, CA) as previously described [122,132,133]. Fluorescence was determined using a Synergy™ 2 Multi-Mode microplate reader (BioTek Instruments, Inc., Winooski, VT).

RNA-Seq

Total RNA was isolated from eight bloodstains as described above. The eight samples used in this analysis included the following: two samples (one male and one female, all Caucasian) from each of four age groups or classifications (pre-adolescent infant/child (ages 1 hour and 33 months), adolescent (ages 12 and 14 years), adult (ages 29 and 31 years) and elderly (ages 71 and 81 years)). The use of male and female donors permitted an assessment of potential affects of gender. Transcriptome sequencing (RNA-Seq) was performed by Beckman Coulter Genomics Inc. (previously Agencourt) using the SOLID™ System. All obtained sequences were mapped to the human RefSeq, hg18 (NCBI). The digital gene expression results included files with raw counts of mapped reads, only genes with a raw count of 100 or more in at least one sample (count.min.100) and normalized counts per million (raw count divided by the total sum of counts per sample times 1,000,000).

cDNA Synthesis (Reverse Transcription, RT)

High Capacity cDNA Reverse Transcription kit (Life Technologies) – for mRNA. The 20μl reaction consisted of: 2μl 10x RT buffer, 0.8μl 25x dNTPs (100mM), 2μl 10X RT random primers, 1μl of MultiScribe™ Reverse transcriptase, 14.2 μl of sample and water (volume of water adjusted accordingly depending on the amount of sample required). The RT reaction was as follows: 25°C for 10 min, 37°C for 120 min, and 85°C for 5 min. The standard RT input for the mRNA profiling assays was 25ng of total RNA. A reverse transcription blank (containing

reverse transcriptase enzyme but no sample) was included for each RT reaction. All RT products were stored at -20°C until needed. A positive control was included in the RT reactions, which consisted of a known blood total RNA sample (previously demonstrated to give consistent and reproducible RNA profiling results). A reverse transcription negative sample (RT-, no reverse transcriptase enzyme added) was also included in each RT reaction. The positive control sample was used to prepare this RT- control.

miScript Reverse Transcription Kit (QIAGEN, Valencia, CA) – for miRNA. The 20µl reaction consisted of: 4µl 5x miScript RT buffer, 1µl miScript Reverse Transcriptase mix, and 15 µl of sample and water (volume of water adjusted accordingly depending on the amount of sample required). The RT reaction was as follows: incubation at 37°C for 60 min, followed by inactivation of reverse transcriptase at 95°C for 5 min. All reactions were performed on an ABI 9700 thermal cycler. The standard RT input for the miRNA profiling assays was 1-5ng of total RNA. A reverse transcription blank (containing reverse transcriptase enzyme but no sample) was included for each RT reaction. Reverse transcription negatives (sample but no reverse transcriptase enzyme) were also included.

miScript Reverse Transcription II Kit (QIAGEN) – for miRNA. The 20µl reaction consisted of: 4µl 5x miScript buffer (HiSpec or HiFlex), 2µl 10x miScript Nucleics mix, 2µl miScript Reverse Transcriptase mix, and 12 µl of sample and water (volume of water adjusted accordingly depending on the amount of sample required). The RT reaction was as follows: incubation at 37°C for 60 min, followed by inactivation of reverse transcriptase at 95°C for 5 min. All reactions were performed on an ABI 9700 thermal cycler. The standard RT input for the miRNA profiling assays was 1-5ng of total RNA. A reverse transcription blank (containing reverse transcriptase enzyme but no sample) was included for each RT reaction. Reverse transcription negatives (sample but no reverse transcriptase enzyme) were included as needed.

Post-RT Purification

For some samples (obtained from Arnold Palmer Hospital), a post-RT purification was required. It is unclear why this purification was required. However without purification, poor or no results were obtained from subsequent real time PCR detection using the *Power*SYBR® reagent (see below for protocol). For purification of these RT reactions, the QIAGEN MinElute purification kit was used (QIACube protocol, 20 µl nuclease free water elution).

Singleplex Pre-Amplification (mRNA)

For some genes, high or undetected Ct values were observed for all samples. For these candidates, we developed an in-house pre-amplification method in an attempt to improve the detection of these genes. The 25 µl reactions consisted of the following: 3 µl cDNA (25 ng total RNA input, see above), 1X PCR Buffer II (10 mM Tris-HCl, pH 8.3, 50 mM KCl) (Life Technologies), 1 mM dNTPs (Life Technologies), 3.25 mM $MgCl_2$ (Life Technologies), 2U AmpliTaq Gold DNA polymerase (Life Technologies), 1.6 µM (final concentration) unlabeled forward and reverse primers (brought to 25 µl with nuclease free water). The cycling conditions consisted of the following: 95°C 11 min; 10 cycles of 94°C 20 sec, 55°C 1 min, 72°C 45 sec;

72°C 5 min. The RT positive control and RT- negative control were also pre-amplified in the same reaction. An amplification blank (nuclease free water) was included with each reaction.

Product Detection – Real Time PCR (mRNA/miRNA)

All Real-time PCR was performed using the Relative Quantitation protocol on an ABI Prism 7000, 7500 or 7500 Fast Sequence Detection System (Life Technologies).

PowerSYBR® (mRNA) – Two microliters of cDNA (or pre-amplified cDNA) was amplified using the *Power*SYBR® reagent (Life Technologies). The 25 µl reaction consisted of the following: 1X PowerSYBR® reagent, 3.2 µM (final concentration) each of the unlabeled F and R primers, volume brought to 25 µl with nuclease free water. All primers were obtained from literature sources or designed using Primer 3. The cycling program consisted of the following: 95°C 10 min, 40 cycles 95°C 15 sec, 60°C 1 min; melt curve analysis (software default program): 95°C 15 sec, 60°C 1 min, 95°C 30 sec, 60°C 15 sec. A positive control (known blood RT+), RT- negative control, DNA (2ng) and an assay blank were included with each real time PCR assay.

MGB Probe – Based Assays (IGFBP3, COL1A2, B2M) – Two microliters of pre-amplified cDNA was amplified in a 25 ml reaction: 1X TaqMan Universal PCR Master mix (Life Technologies), 1.6 µM primers (unlabeled) (final concentration), 0.2 µM MGB probe (final concentration) (reaction brought to 25 µl with nuclease free water). The cycling program consisted of the following: 95°C 10 min; 45 cycles: 95°C 15 sec, 60°C 1 min. A positive control (known blood RT+), RT- negative control, DNA (2ng) and an assay blank were included with each real time PCR assay.

miScript SYBR Green PCR kit - One microliter of cDNA was amplified using the miScript SYBR® Green PCR kit (QIAGEN) (25µl reaction: 12.5µl 2x QuantiTect SYBR Green PCR master mix, 2.5µl 10x miScript universal primer (1X final concentration), 2.5µl 10x miScript primer assay (1X final concentration), according to the manufacturer's protocols with one minor change (cycle number modified to 45 for initial screening and assay development in order to evaluate possible low abundance candidates). All miScript primer assays were obtained from QIAGEN. The cycling program consisted of the following: 95°C 15 min, 40-45 cycles of 94°C 15 sec, 55°C 30 sec, 70°C 34 sec. The time for the 72°C extension step was increased to 34 sec (from 30 sec) since the 7500 real instrument requires a minimum of 34 sec (detection step). While only required for the 7500 real time instrument, for consistency in programs between instruments 34 sec was used for all instruments. A positive control was included for all real time PCR reactions, which consisted of a known blood sample RT product that had been tested previously (positive result obtained). An assay blank, consisting of all reaction mix components except that nuclease free water was used in place of sample, was included for each real time PCR reaction.

miScript miRNome miScript miRNA PCR Array (V16.0) – For the miScript miRNA PCR arrays, the recommended RNA starting amount was ~250-500 ng per four 96 well plates (12 plates per array, so x3 reactions needed). For this analysis, one adolescent (15 years old, female)

and one elderly (89 years old, male) sample were selected. For the elderly sample, ~255 ng (x3) was used in the miScript reverse transcription reaction (HiSpec buffer). For the adolescent sample, ~286 ng (x3) total RNA input was used in the miScript reverse transcription reaction (HiSpec buffer). Each RT was diluted prior to use in the miScript SYBR green real time PCR assays (20 µl RT reaction, 90 µl nuclease free water). The miScript SYBR® Green PCR kit (QIAGEN) reaction consisted of the following: 1375 µl 2x QuantiTect SYBR Green PCR master mix, 275 µl 10x miScript universal primer, 25 µl template cDNA (providing 0.5-1 ng cDNA per well) and 1075 µl nuclease free water, according to the manufacturer's protocols. Twenty-five microliters of reaction mix was dispensed into all wells of the 96-well array plates. The cycling program consisted of the following: 95°C 15 min, 40 cycles of 94°C 15 sec, 55°C 30 sec, 70°C 34 sec. All plates were run on an ABI 7500 Fast real time PCR instrument. Positive and negative controls were included on each array plate and will be described in the results section.

III. RESULTS

A. Statement of Results

1. Approaches to candidate searches and identification

In this work, we sought to identify age-specific biomarkers with the ultimate aim of designing assays that may be used to predict the biological age of a sample donor. An individual's age is innately associated with genetic changes. Several metabolic and biochemical processes within the body are key to specific developmental stages in a person's life and these are regulated by specific genes. Thus the possibility exists of finding an association between patterns of mRNA expression of such regulatory genes, and the genes that they regulate, and specific stages of life. Below we provide a general overview of the approaches to biomarker identification used throughout the study as well as a description of the samples used in the majority of the studies described.

Candidate biomarker identification

We took a two-pronged approach to mRNA biomarker identification throughout this project. The first strategy included the analysis of entire transcriptomes of selected samples of different ages, using deep sequencing mRNA-Seq technology, in order to identify genes that show different expression with respect to the age of the donor. The second approach involved targeting specific candidate genes that are, or are likely to be, involved in regulation of age-related processes based upon *a priori* understanding of the physiology and biochemistry of human development. Both approaches were ultimately successful in the identification of potential age biomarkers.

In order to carry out the RNA-Seq experiments, we used 8 different samples (all Caucasian) of various ages (pre-adolescent children, adolescent, adult and elderly age groups). The precise samples and ages were as follows: 1 hour (female), 33 month (male), 12 year (female), 14 year (male), 29 year (male), 31 year (female), 71 year (female) and 81 year (male). In order to try and account for any potential differences due to gender, we chose one male and one female sample from each of the four different age groups. Total RNA was extracted from each of these samples using an organic (acid guanidinium thiocyanate-phenol-chloroform) extraction protocol. These total RNA extracts were sent to Beckman Coulter Genomics Inc. for next generation sequencing (NGS) using the SOLiD™ System (Applied Biosystems). This particular platform enables massively parallel sequencing of clonally amplified DNA fragments linked to beads.

From the RNA-Seq analysis, we obtained digital gene expression (DGE) results for all detected genes. All obtained sequences were mapped to human RefSeq, hg18 (NCBI). The mapping results provided details such as the number of mapped reads, the number of uniquely mapped reads, and genome coverage (100% coverage - % not covered; high value is good). The DGE results included files with raw counts of mapped reads, genes with a raw count of 100 or more in at least one sample (count.min.100 file), and normalized counts per million (raw count divided by the total sum of counts per sample times 1,000,000). Links to Entrez Gene ID and description of identified genes were also provided. Over 40,000 genes were identified in the samples. Table 1 contains a representative example of the DGE results for each of the samples.

An evaluation of the count values for each gene provided an indication of the relative abundance amongst the various age groups tested and permitted an identification of potential candidates for each age group. Initially, it was assumed that the genes with the highest count values in each age group would be the best potential age biomarkers. However, initially it was somewhat unclear how the count number correlated with detectible expression as detectable on gel or qPCR based assays. We have extensive previous work experience with the use of mRNA profiling for body fluid identification, with numerous mRNA biomarkers for the identification of blood. We therefore mined the sequencing results in order to determine if these genes were present (quality control check) and to evaluate the count values for these genes in order to try to estimate the relative abundance levels of other genes based on the data from these genes for reference. Most of our routinely used blood markers were found in the sequencing data. However, the count number for these genes were relatively low (<1,000) except for HBB and HBA. These genes are readily detected in total RNA from dried bloodstains and are therefore present in sufficient quantities for analysis. The genes from the NGS analysis with significantly higher count numbers than our body fluid ID biomarkers must be highly abundant. As a result we evaluated candidates with significantly high-count numbers that demonstrated differential expression between the age groups as well as candidates with lower overall count values but still showing differential expression amongst the various age groups.

Using both approaches, we screened over 500 mRNA biomarkers (Table 2). Approximately 20% of the candidates were identified from RNA-Seq data with the remainder being obtained using the literature/physiological function based candidate gene approach.

Samples used in gene expression profiling experiments

For all studies, we examined gene expression in samples of various ages and stages of life, from newborn to elderly. Initial sample age classifications were as follows: newborn/infant (< 2 years), toddler (2-6 years), child (7-12 years), adolescent (13-17 years), adult (18-45 years), mature adult (46-60 years) and elderly (>60 years). These initial age sub-categories represent the developmental phases of the human life cycle and were used heuristically in the event that distinct biomarkers for each age group could be identified. However, we realize that these designations may not necessarily represent absolutely distinct 'gene expression boundaries' and the precise ages that differentiate the age groups may be different to the ones we chose on an ad hoc basis. Therefore, for interpretation of any obtained expression data throughout this project we were careful not to rely strictly on these classifications and to evaluate the data in a broader non-biased sense without pre-judging to which category any sample (that was close to an age classification boundary) belonged to.

Our screening strategy employed and initial sample screening set that included multiple donors of each of the age categories in order to provide an initial indication of gene expression in a wide range of age groups. The screening sample set typically included twenty-one samples (both male and female samples): 3 newborn/infant, three toddler, 3 child, 3 adolescent, 4 adult, 2 mature adult and 3 elderly samples. Within each age group, we tried to vary the ages to span that particular age range. If an expression trend was identified for a particular candidate using these screening set, we performed additional analyses with a confirmation sample set which again included multiple donors (all different from those contained in the screening set) for each of the age categories in order to provide additional data points within each age category. A typical confirmation set included the following: 6 newborn/infant, 3 toddler, 2 child, 2 adolescent, 5

adult, 3 mature adult and 5 elderly samples. Additional donors within each age classification were available as well for any assays requiring additional donors for further validation.

2. mRNA

Table 2 contains a list of 538 mRNA gene candidates that were evaluated as potential age biomarkers. Once suitable primers were designed, the expression of each candidate was evaluated using the screening sample set (described in the previous section). Complementary DNA (cDNA) was produced from the total RNA isolates using reverse transcription. Firstly the cDNA samples were amplified with a set of housekeeping gene primers (B2M) to ensure consistent expression levels among the samples. Real time PCR detection was mainly employed for detection of PCR product since it proved to be difficult to assess subtle expression differences using gel electrophoresis. There was also no easy way to normalize expression data using gel electrophoresis. For real time PCR analysis, the design and use of fluorescently labeled probes would have been quite costly. Therefore, we utilized a SYBR green based method for product detection using real time PCR (*Power*SYBR™ reagent, Life Technologies). The *Power*SYBR™ assays required only unlabeled primers rather than expensive fluorescently labeled probes. The use of the *Power*SYBR™ reagent permitted the use of melt curve analysis on individual candidates. This allowed us to determine if different products were potentially being detected in different age groups. This was quite valuable for the identification of possible isoforms of individual genes expressed at different stages in the life cycle.

The expression level of each candidate was evaluated in the initial screening sample set. Delta-Ct (dCt) values were calculated upon normalization with B2M. The dCt values were then compared between the various age groups (newborn-infant, toddler, child, adolescent, adult, mature adult, elderly) in order to determine if an upwards or downwards trend in expression in relation to biological age could be identified. A candidate was selected for further analysis (testing of additional donors in each age group) if a trend was identified. While numerous candidates were selected for further testing, a majority of candidates were subsequently rejected upon further testing since the originally observed expression trend was no longer observed with additional samples. This screening strategy allowed us to concentrate on the most promising candidates and, as a result, we successfully identified numerous gene candidates that demonstrated promise as potential age biomarkers. Summaries of those gene candidates are provided below.

COL1A2/IGFBP3

In previous work, we obtained preliminary data that indicated the potential ability to provide low-resolution differentiation of four possible age groups using COL1A2 (associated with bone development) and IGFBP3 (a regulator of somatic growth and cellular proliferation) (see Figure S1, Literature Review). While the use of these two candidates alone was not sufficient to permit a clear differentiation of all age groups, there was a some clustering of samples < 5 years old and > 5 years old. Within each of these groups, although again not completely separated, there was a further sub-classification of ages (e.g. 5-18 years vs. > 18 years). Therefore, we hypothesized that, at the very least, these candidates might serve as an age "milestone" that could be useful in a broader older/younger age classification.

The observed Ct values for both COL1A2 and IGFBP3 in previous testing as well as initial additional testing in the current study were relatively high ($\geq$ 33 Ct) indicating low gene exression. Even with additional input total RNA (up to 50 ng), relatively high Ct values were still obtained. In an attempt to improve sensitivity prior to qPCR detection, we developed an in-house strategy using a 10-cycle pre-amplification reaction with the same PCR primers used for subsequent qPCR detection. This simple amplification step prior to qPCR detection resulted in a significant improvement in the observed Ct values for COL1A2 and IGFBP3 (~27 – 30 Ct values). For COL1A2, lower Ct values were observed for younger samples (newborn – child) while lower Ct values were observed for older samples (child – elderly) for IGFBP3. The dCt values for COL1A2 and IGFBP3 are shown in Table 3. Previously, our analysis of such dCt data would include the development of a 2D scatterplot assay using the normalized expression data from these two candidates. However in separate work involving miRNA profiling assays for the identification of forensically relevant biological fluids, we developed novel logistic regression (LogR) statistical models that provided accurate predictions of the presence of forensically relevant fluids [134]. With the success of these models in predicting the body fluid of origin of dried biological stains, we decided to attempt to develop LogR models based upon our age classification assays.

With the assistance of a statistician at the University of Missouri Kansas City (UMKC) (Dr. Kamel Rekab), a binary logistic regression (LogR) statistical model was developed using the COL1A2/IGFBP3 assay that permitted an identification of newborn/toddler samples. All data analyses were performed using the statistical software SPSS version 19. The data set consisted of 67 donors (Table 3). 53 were classified as "0" (child or older samples) and 154 were classified as "1" (newborn/toddler samples). The goal of an analysis using logistic regression is the same as any modeling technique in statistics: to find the best fitted model to describe the relationship between an outcome (response) and a set of particular variables. What distinguishes a logistic regression model from the linear regression model is that the outcome is binary. The log distribution (or logistic transformation of p) is also called the logit of p or logit(p). It is the log (to the base e) of the odds ratio or likelihood ratio that the outcome is a newborn-toddler sample. In symbols it is defined as $\text{logit}(p) = \ln[p/(1-p)]$, where p denotes the probability of identifying a newborn-toddler age correctly. The estimated logit based on 67 selected cases is given by the following expression: $\ln[p/(1-p)] = -0.071$ (IGFBP3*COL1A2) + 2.480 (IGFBP3) + 0.048 (COL1A2^2) (the gene names in the equation represent the normalized expression data compared to B2M for each gene).

In the model, the chi-square has 3 degrees of freedom, a value of 58.800 and a probability of $p < 0.000$. This indicates that the COL1A2 and IGFBP3 and their interaction have a very significant effect. The Nagelkerke R square indicates that 91.1% of the variation in the newborn-toddler data is explained by the logistic model. This shows that there is a strong relationship of 91.1% between all the three predictors (IGFBP3, COL1A2 and the two-factor interaction) in the prediction equation. Furthermore, the Hosmer and Lemeshow (H-L) test shows that the model prediction does not significantly differ from the observed. A probability (p) value is computed from the chi-square distribution with 8 degrees of freedom to test the fit of the logistic model. Since the H-L goodness-of-fit statistic is 0.974, which is greater than 0.05, which we want for well fitting models, we fail to reject the null hypothesis that there is no difference between observed and model-predicted values thus implying the model's estimates fit the data at an acceptable level. The receiver-operating characteristic (ROC) curve is very useful in evaluating the predictive accuracy of a chosen model in logistic regression. We obtain the curve by the

predictive accuracy of a chosen model in logistic regression. We obtain the curve by plotting sensitivity (True Positive, TP) against 1-specificity (False Positive, FP). A perfect classification with 100% true positive and 0% false positive would have an area equal to 1. For this analysis, the area under the curve is 99.5% with a 95% confidence interval (98.2%, 100%). The optimal threshold was determined as the threshold corresponding to the nearest (FP,TP) to (0,1). In this analysis, the optimal threshold was 0.45.

Once the logit is estimated, then the probability that the outcome is that the donor belongs to the newborn/toddler age group is given by: $P=\exp(\ln(O))/[1+\exp(\ln(O))]$, where the odds ratio $O = p/(1-p)$. Since the optimal threshold is 0.45, then $P > 0.45$ will be classified as a newborn/toddler sample and $P \leq 0.45$ will be classified as a non-newborn-toddler sample. Using this model 52/53 non-newborn/toddler samples (i.e. child or older: 10 child, 9 adolescent, 17 adult, 6 mature adult and 11 elderly samples) were correctly classified as non-newborn/toddler and 14/14 of the newborn/toddler samples (9 newborns and 5 toddlers) were correctly classified as newborn/toddler. There was one false positive, an 8 year-old sample, which was incorrectly classified as a newborn/toddler sample. However, toddler samples in our study included ages 2 – 6 years. Therefore, the 8 year-old sample is not far from this somewhat arbitrarily set age range and therefore may not be as significant a false positive as would an elderly sample giving the same result for example. Further validation of this assay will be needed in order to more clearly delineate the age range boundaries for the newborn/toddler age classification. However, overall the results of the LogR models suggest that this is a useful assay in predicting a bloodstain donor as being of newborn/toddler age (or not). Additionally, there does not appear to be a gender affect with this assay. Each age group includes both male and female samples. For example, the newborn samples samples include 4 male and 5 female samples and the toddler samples include 3 female and 2 male samples. Since all of the 14 newborn-toddler samples were correctly identified, there is no indication with this assay that gender has a significant effect on gene expression resulting in a mis-classification of either male or female samples.

The results of the COL1A2/IGFBP3 assay permit a newborn/toddler versus all others age classification. In our previous work, we were only able to identify newborns (using different gene biomarkers) versus all others. Therefore, the COL1A2/IGFBP3 model extends our age predictive ability beyond just newborns to also encompass the toddler age group (2-6 yrs).

Elderly

The COL1A2 /IGFBP3 assay described above resulted in a prediction model for a newborn/toddler age classification. We subsequently identified 10 potential mRNA candidates whose expression appeared to result in clustering of elderly samples when used in various 2D plot combinations. The ten candidates include: C1QB, C3, GFI1, GSTM3, UQCRQ, IGJ, UQCRQFS1, SENP1, FCGR2B and NDUFB11. The elderly samples were not sufficient separated from all other ages on any of the individual 2D plots (a sub-set of the 2D plots are shown in Figure 1). However, the use of logistic regression modeling permits the development of multi-candidate models instead of the more limiting two-candidate scatter plots. Therefore, we again used logistic regression modeling to evaluate the 10 potential elderly candidates to try to develop a model that would permit an identification of elderly samples.

We were able to successfully develop a LogR model for the prediction of elderly samples using five of the ten potential gene candidates. The estimated logit based on 46 selected cases is given by the following expression: $\ln[p/(1-p)] = -5.563(IGJ)+1.620(UCQCRFS1)-$

2.652(SENP1)+0.494(FCGR2B*IGJ)-0.180(FCGR2B*NDUFB11) (the gene names in the equation represent B2M-normalized expression values).

In this model, the chi-square has 5 degrees of freedom, a value of 28.990 and a probability of $p < 0.000$. This indicates that the gene and interaction factors in the model have a very significant effect. The Nagelkerke R square indicates that 76.9% of the variation in the newborn-toddler data is explained by the logistic model. This shows that there is a strong relationship of 76.9% between all the predictors in the prediction equation. For this analysis, the area under the curve is 95.8% with a 95% confidence interval (87.8%, 100%). The optimal threshold was 0.50.

Once the logit is estimated, then the probability that the outcome is a newborn-toddler age is given by: P=exp (ln(O))/[1+exp(ln(O))], where the odd ratio O = p/(1-p). Since the optimal threshold is 0.500, then $P > 0.500$ will be classified as an elderly sample and $P \leq 0.500$ will be classified as a non-elderly sample. Using this model 39/39 non-elderly samples (9 newborn, 6 toddler, 5 child, 5 adolescent, 9 adult and 5 mature adult) were correctly classified as non-elderly and 7/8 of the elderly samples were correctly classified as elderly. There was one false negative result, in which a 66 year-old sample was incorrectly classified as non-elderly. This was the 'youngest' elderly sample (ages 66, 71, 73, 76, 81, 84, 89 and 91 years) and closer to the mature adult age range (46 – 60 years). Therefore, the exact boundary for the elderly samples using this assay may possibly be > 70 years instead of 65 years defined by our original age classifications. We will need to evaluate addition samples between 60-70 years of age in order to determine where the "elderly" classification truly lies. However, there is a plausible explanation for this false negative result and it therefore does not represent a significant weakness of the assay. Additionally, there does not appear to be a gender affect with this assay. Each age group includes both male and female samples. For example, the elderly samples include 4 male and 4 female samples. Since seven of the eight samples were correctly identified, there is no evidence that gender has a significant effect on gene expression resulting in a misclassification of either male or female samples.

As with the previous COL1A2/IBFBP3, additional validation work will be required on each of these assays in order to optimize and finalize the assays before they can be used routinely for age prediction. Despite the need for additional validation, the results of the elderly assay are promising and represent the ability to distinguish another age group.

Young vs older

The two above described assays permitted an identification of toddler and elderly samples (in addition to our previous newborn assay) thus representing three of the seven original age classifications originally defined in this study. During the course of the evaluation of the elderly candidates, we also identified six additional candidates (with some overlap with the previously evaluated elderly candidates) that possibly permitted a distinction between younger (newborn through child ages) and older (adult through elderly) samples. The six candidates included the following; CCR7, RBBP8-v3, GFI1, FCGR2B, IGJ and GPN1. However, we were also able to consider the previously described elderly candidates in the development of this assay as these gene candidates were evaluated at the same time and using the same sample set. Upon evaluation of these candidates, a logR model was developed that permits and distinction of younger (newborn-child) vs older samples (adult-elderly). For initial construction of this assay, the adolescent samples were not included (42 samples without the 5 adolescent samples) since it was unclear whether they would be considered in the younger or older age groups. We are currently

in the process of evaluating the model with adolescent samples that may result in minor modifications to the existing assay. However, we are confident that the model will still permit a distinction between younger and older samples. The only difference will be whether adolescent samples are classified as older or younger. This will be discussed in more detail in the subsequent section.

For this assay, the estimated logit based on the 42 selected cases is given by the following expression: ln[p/(1-p)] = 4.397(CCR7) - 1.143(RBBP8-v3) – 1.801(GFI1) + 0.468(GSTM3) - 1.391(SENP1) (the gene names in the equation represent the B2M normalized expression data No interaction terms were used in this model. This model is constructed so that the classifications are older ("1") or younger (or non-older, "0").

In this model, the chi-square has 5 degrees of freedom, a value of 45.541 and a probability of $p < 0.000$. This indicates that the gene factors in the model have a very significant effect. The Nagelkerke R square indicates that 88.3% of the variation in the older sample data is explained by the logistic model. This shows that there is a strong relationship of 88.3% between all the predictors in the prediction equation. The optimal threshold was 0.500.

Once the logit is estimated, then the probability that the outcome is an 'older' age sample is given by: P=exp (ln(O))/[1+exp(ln(O))], where the odd ratio O = p/(1-p). Since the optimal threshold is 0.500, $P > 0.500$ will be classified as an 'older' sample (adult-elderly) and $P \leq 0.500$ will be classified as a younger (or non-older) sample (newborn – child). Using this model 21/22 older samples (9 adult, 5 mature adult and 8 elderly samples) were correctly classified as older samples (1 false negative) and 19/20 younger samples (9 newborn, 6 toddler and 5 child) were correctly classified as younger (or non-older) (1 false positive). We are currently evaluating the data to evaluate the false negative and false positive result in order to determine the nature of these misclassifications.

Additional validation work is needed for this model, including the assessment of adolescent samples. Additional samples for each age group will need to be run to fully evaluate the model. However, the initial results from this model are promising for a younger/older age classification.

Summary – mRNA profiling assays

Using the above-described assays in combination with our previously developed newborn identification assay, we have the potential to classify the biological age of a bloodstain donor into one of five age classifications. Figures 2 and 3 contain classification flow charts for unknown bloodstains, each resulting in five distinct age classifications. Since we are still finalizing the younger vs older assay (evaluation of adolescent samples currently being performed), there are two potential pathways that the assays can follow depending on whether the adolescent samples are classified as 'younger' or 'older'. Each will be described below.

Figure 2 depicts the workflow assuming that the adolescent samples are classified as 'younger' in the younger/older assay. For an unknown sample using this workflow, samples would first be analyzed using the younger/older assay. This assay will classify a sample as 'newborn/toddler/child/adolescent' (younger, $P \leq 0.5$ using the developed logR model) and 'adult/mature adult/elderly' (older, $P > 0.5$ using the developed logR model). These are two distinct groups and in some cases may provide sufficient age classification on its own depending on the circumstances of criminal investigations. If the sample was classified 'older', it would then be tested using the elderly (IGJ/UQCRFS1/SENP1/FCGR2B/NDUFB11) assay, which will further classify these samples into elderly ($P > 0.5$) and adult/mature adult ($P \leq 0.5$). At this time,

no further classification of these samples would be necessary as an adult or elderly classification will be made. It is unclear at this time whether suitable candidates can be identified to distinguish adult and mature adult samples and therefore a single adult classification would be required. With further validation of the developed assays and logR models, we will optimize the included ages in each category. If the sample was classified as 'younger', it would then be tested using the newborn-toddler (IGFBP3/COL1A2) assay, which would further classify these samples into 'newborn/toddler' ($P < 0.450$) and 'child/adolescent' ($P > 0.450$). The 'newborn/toddler' group could then be further classified into 'newborn' and 'infant-toddler' using our previously developed newborn (HBG1n/HBG2n) assay. While a logR model was not originally developed for this assay since the classification was based upon two genes and the gene expression clustering between newborn and non-newborns was non-overlapping, we will develop one in subsequent work. Figure 3 depicts the workflow assuming the adolescent samples are classified as 'older' in the younger/older assay. Using this alternative workflow, only minor differences to the age group classifications are necessary. The same progression of assays would be performed for this workflow. Newborn, infant-toddler and elderly classifications remain unchanged. However new classifications of child and adolescent/adult/mature adult are made as a result of adolescent samples classified in an 'older' sample group compared to the first assay. We believe that the adolescent samples will be classified as "younger" in the logR models so that a child-adolescent vs adult classification can be made. Alternatively, an adolescent through mature adult age classification is quite a broad spectrum of ages and so we are hopeful that a more refined age classification is possible as wouod be the case with a child/adolescent classification. However, for both workflows, the identification of an adolescent candidate assay would provide the necessary refinement of the child/adolescent or adolescent/adult/mature adult groups. We remain hopeful based on the results of the existing candidates and assays that an adolescent classification will be possible.

Using either of these workflows (~15 genes: CCR7, RBBP8-v3, GFI1, GSTM3, SENP1, IGFBP3, COL1A2, IGJ, UQCRFS1, FCGR2B, NDUFB11, HBG1n, HBG2n, B2M, S15; Table 4), age classifications of 1) newborn, infant/toddler, child/adolescent, adult/mature adult and elderly or 2) newborn, infant/toddler, child, adolescent/adult/mature adult and elderly can be made. Additionally, currently expression data from a majority of these candidates are obtained using *Power*SYBR™ assays and do not require the use of fluorescently labeled probes thus reducing the overall cost of analysis. The results from this workflow provide the ability to distinguish a significant number of the original age classifications that we hoped to achieve at the beginning of the study.

3. miRNA

miScript qPCR – Individual Assays

While the original experimental design of the project involved the analysis of mRNA, in separate work we have had great success with miRNA profiling assays for body fluid identification. We conducted an initial comprehensive screening of 452 miRNAs in an attempt to identify body fluid-specific miRNAs. We successfully identified a panel of nine differentially expressed miRNAs that permit the identification of blood, semen, saliva, vaginal secretions, menstrual blood and skin. Additionally, we successfully used logR models for an accurate and definitive determination of the body fluid of origin. During the development of these assays, a

significant number of miRNAs were found to be highly expressed in blood. Due to the success of the miRNA assays, the large number of highly abundant miRNAs in blood, and improved sensitivity compared to mRNA profiling (miRNA assays require as little as 25 pg instead of 25 ng for detection), we hypothesized that it might also be possible to identify miRNA biomarkers that show differential expression in blood as people age.

Currently, we use the commercial Qiagen miScript system for miRNA expression profiling. The miRNAs are simultaneously polyadenylated and reverse transcribed using both random and oligo-dT primers. The oligo-dT primers contain a 5' end universal tag, which is utilized as a primer binding site for subsequent real time PCR detection. Unlike some other commercially available miRNA systems that employ miRNA-specific reverse transcription strategies, the miScript system allows cDNA to be produced from all RNA species within the sample, including mRNAs and other small non-coding RNAs. This could be advantageous in forensic casework whereby multiple individual reverse transcription reactions required to analyze numerous miRNAs using other standard approaches may be impossible due to sample limitations. The ability to simultaneously reverse transcribe other RNA species, such as mRNAs, may also be useful if parallel assays need to be performed (tissue specific mRNAs for simultaneous confirmation of the blood origin of the sample, housekeeping genes for normalization or mRNA-age candidates etc). Individual miRNAs are subsequently detected using the miScript SYBR® Green PCR kit. Each single-plex reaction utilizes a universal primer (complementary to the 5' universal tag on oligo-dT primer used in the reverse transcription reaction) and a miRNA-specific primer.

Using the miScript system, we evaluated the expression of 1,186 miRNAs (current through v.16 of miRBase) in bloodstains from donors of varying ages representing a majority of the defined age categories: newborn (1 hour, male), toddler (2 year, female), adolescent (13 year, male), adult (32 year, female), mature adult (58 year female) and elderly (73 year male). Only one donor for each group was used due to the large number of miRNA assays. One nanogram of input total RNA was used for analysis. Raw Ct values of all 1,186 miRNAs were obtained and plotted in simple line graphs in order to try to identify potential.

From the original screening of the 1,186 assays, possible miRNA biomarkers for each age group were identified: 34 infant, 3 child, 15 adolescent, 8 adult, 3 mature adult, and 10 elderly. These candidates demonstrated increased expression in these single age categories. However, there were additional miRNAs that demonstrated increased expression in broader age ranges: infant through adolescent (6), infant through child (3), child through adolescent (2), child through adult (2), adult through elderly (11) and mature adult through elderly (2). In order to detect possible lower abundance candidates, 5 ng input total RNA was also tested. With this increased input amount, several additional candidates were identified: 9 adult, 6 mature adult, 30 elderly, 17 mature adult through elderly, and 8 adult through elderly. Overall, there were ~170 candidates that were identified from the initial screening data.

Additional samples were tested for each age group. For the additional testing, a set of 21 samples ranging in age from 1 hour to 93 years was used. The expression of each of the ~170 miRNA candidates was evaluated using this sample set. The expression data from the additional sample set was then evaluated to determine if the identified trend was reproducible. The trend observed with the initial screening data was not reproducible in the larger sample set.

Subsequent to this initial testing, we had begun work on the logR models for both body fluid identification (separate work) and age assays. The use of logistic regression modified how we assessed potential RNA biomarkers since multi-factor (i.e multi-gene) gene models were

possible and the logR models could also include interaction factors that would not be obvious from a visual examination of raw or dCt values. Therefore, after the initial analysis was not particularly fruitful, we re-evaluated our existing expression data to identify potential candidates that might have been previously overlooked. When the expression data was re-evaluated, we identified 84 additional potential miRNA candidates. The identified miRNA candidates for various age groups are provided in Table 5. For the additional testing of these candidates, we normalized the expression data using miR-940 (the normalized used in our miRNA body fluid identification assays). The expression level of each candidate was evaluated first using the screening sample set (described previously). Any candidate demonstrating a possible age trend was further evaluated using additional samples (confirmation set). Twenty-one of the 84 candidates were selected for additional testing: miR-30e, miR302a, miR-588, miR-3196, miR-3921, miR-1284, miR-1251, miR-888, miR-513c, miR-1827, miR-621, miR-598, miR-517b, miR-337-5p, miR-34b, miR-941, miR-384, miR-523, miR-516a-5p, miR-520a*, miR-525. Every possible 2D plot was evaluated using each of these candidates. Twenty-five 2D plot combinations from among these 21 candidates demonstrated potential trends with age, mainly older/younger classifications. An example of one of the 2D plots is shown in Figure 4. While none of the 2D scatterplots on their own are sufficient to distinguish individual age groups, we are hopeful that with logR models and the potential to combine more than two candidates in a single model that a suitable assay can be developed. We are currently working to analyze the data using logistic regression analysis in order to develop potential models.

miScript miRNome miRNA PCR Arrays

Despite the successful identification of 21 potential age biomarkers using the above described approach of screening 1,186 individual miScript primer assays, the individual assay screening approach has certain limitations. For the initial screening, we were only using one donor per age group and therefore it is often difficult to identify or confirm potential expression trends based on individual samples. Candidate selection from the initial screening data was also based on raw Ct values and not from normalized expression data. While this approach can still result in the identification of suitable candidates after further testing is performed (this is the same approached used to identify candidates for our miRNA body fluid identification assays and was largely successful for that application), we decided to utilize miRNA PCR arrays in order to further evaluate a large number of miRNAs. The QIAGEN miScript® miRNA PCR Array plates provide the opportunity to evaluate expression levels of 1,008 of the most abundantly expressed and best characterized miRNA sequences in the human miRNA genome based on data in miRbase v.16.0. There were several advantages to using the miScript® array plates: 1) it is the same chemistry and platform that we would use for subsequent additional testing thereby removing any potential expression differences observed when different platforms or technologies are used [122,135], 2) a set of controls is included on each plate which allows the use of a ΔΔCt data analysis approach for relative quantification, assessment of reverse transcriptome performance and assessment of PCR performance, and 3) free and user-friendly data analysis software is available online for data analysis of the obtained expression data. The controls included on each array plate included the following: 1) C. elegans miR-39 (for possible alternative data normalization using an exogenously spiked Syn-cel-miR-39 miScript miRNA mimic), 2) six snoRNA/snRNA (small nucleolar and small nuclear RNA, respectively) miScript primer assays: SNORD61, SNORD68, SNORD72, SNORD95, SNORD96A, RNU6-2 (for data

normalization using the ΔΔCt method for relative quantification), 3) miRTC (reverse transcription control) (for assessment of reverse transcription performance and 4) PPC (primer assay positive control) (for assessment of PCR performance). For this analysis, we selected one adolescent (15 year, female) and one elderly sample (89 year, male). At the time of analysis these were two target age groups that had not been resolved using mRNA analysis (although subsequently mRNA assays were developed to permit an identification of elderly samples) and represented significantly different stages of the human life cycle.

Expression data for all 1,008 miRNAs was obtained for both the adolescent and elderly sample. For data analysis purposes, one group needed to be selected as the control group. In this study, the adolescent sample was selected as the 'control group' and the elderly sample was selected as 'group 1' for comparison. All relative expression experiments are therefore based on differences in the elderly sample (over or under expression) compared to the adolescent sample. Figures 5 and 6 provide the overall data summary for each sample (adolescent and elderly, respectively). For the adolescent (15 year old) sample, undetectable or absent calls were observed for a significant percentage of the miRNAs tested (~71%). While this was a significant number, 300 miRNAs with detectable and sufficient Ct values were available for analysis (Cts ranging from <25 to 35; anything over a Ct of 35 was considered a negative or absent result). Due to the often limited amount of biological material present in forensic samples, we are looking for highly abundant candidates and therefore we were confident that suitable potential candidates could be identified in the 330 miRNAs (note only ~170 candidates from the 1,186 individual primer assays were selected for further analysis so the array results were consistent with our previous findings). Similar results were observed for the elderly sample (89 year old) with absent calls for ~82% of the miRNAs tested with 205 potential miRNA candidates with detectable and sufficient Ct values.

The next step in data analysis was to select suitable normalizers from the included snRNA/snoRNA controls. The Ct values of each control was compared between adolescent and elderly samples with the goal of selecting only those controls that demonstrated suitable ($Ct \leq \sim 30$) and consistent Ct values in both samples across all array plates. From the six controls, only SNORD95 demonstrated sufficient and consistent Ct values in both samples and on all array plates. Therefore SNORD95 was selected for normalization of the expression data.

Using the available online software, we were able to analyze the expression data in various formats, including scatter plots (Figure 7) and expression heat maps (data not shown), in order to identify potential miRNA candidates with significant changes in expression levels between the adolescent and elderly samples. The software provided fold change values and also provided classifications for the candidates: A - average threshold cycle is relatively high (> 30) in either the control or the test sample, and is reasonably low in the other sample (< 30) (these data mean that the miRNA's expression is relatively low in one sample and reasonably well detected in the other sample suggesting that the actual fold-change value is at least as large as the calculated and reported fold-change result); B - average threshold cycle is relatively high (> 30), meaning that its relative expression level is low, in both control and test samples, and the p-value for the fold-change is either unavailable or relatively high ($p > 0.05$) (this fold-change result may also have greater variations; therefore, it is important to have a sufficient number of biological replicates to validate the result for this gene); C - average threshold cycle is either not determined or greater than the defined cut-off value (default 35), in both samples meaning that its expression was undetected, making this fold-change result erroneous and un-interpretable. From this assessment, we focused our efforts on the "A" classified samples (37 miRNAs, Table 6). For this

analysis, we continued to use SNORD95 for normalization since it was used in the original data analysis. We assessed the relative stability of SNORD95 expression in the sample sets used for this additional analysis and the average Ct values for SNORD95 were relatively consistent amongst all age groups: newborn – 24.6, toddler – 26.7, child – 27.7, adolescent – 26.6, adult – 24.2, mature adult 23.8 and elderly 24.7.

For each of the A candidates, we further evaluated expression levels in an expanded sample set so that multiple donors from the various age groups would be evaluated: 3 toddler, 2 child, 2 adolescent, 5 adult, 3 mature adult and 5 elderly. With this data, we constructed every possible combination of 2D scatterplot and evaluated each assay for the identification of potential age candidates. From the original 37 candidates, we selected 15 miRNAs (miR-421, miR-324-3p, miR-4301, miR-363-3p, miR-221-3p, miR-378a-3p, miR-629, miR-1260a, miR-4286, miR-19b-3p, miR22-3p, miR-29c-3p, miR-19a-3p, miR-151a-3p and miR-3607-3p) for additional testing. An example 2D scatterplot using two of these selected candidates (miR-221 and miR-1260) is shown in Figure 8. As can be seen from this initial analysis, it appears that toddler, child and possibly adolescent samples are found in a distinct cluster, separate from the older samples (adult through elderly). While this is not a perfect separation of samples, since there is a sufficient indication of a possible trend in expression in relation to the age of the donor, we selected these candidates for further analysis as they can be combined with other suitable candidates in the development of the logR models to improve age predictability. This initial screening involves only a small number of samples and therefore all selected miRNAs need to be evaluated with additional donors for each age group. We are currently working on increasing the number of donors for each of these miRNAs to include the following total number of donors: 6 newborn, 9 toddler, 20 child, 10 adolescent, 10 adult, 5 mature adult and 8 elderly. At this time, we are still currently obtaining and reviewing the additional data for each of these miRNA candidates. We are hopeful that we can develop suitable logistic regression models using these miRNA candidates, in possible combination with the additional candidates identified in the individual primer assay testing, to permit an identification of the biological age of a bloodstain donor. It is possible that these assays may be able to further refine the classifications made by the mRNA profiling assays. We are hopeful that the miRNA candidates may possibly provide additional discrimination for adolescent or adult/mature adult samples that is needed to complement the mRNA analysis and provide a full classification of all of the original age classifications established at the beginning of this work.

Quantitative Nuclease Protection Assays (qNPA)

In addition to an evaluation of miRNAs using the miScript assays (individual primer assays and array plates (QIAGEN miScript system), we also investigated an alternative approach to miRNA expression profiling. The alternative strategy involved the use of qNPA (quantitative nuclease protection assay) technology to evaluate gene expression (HT Genomics, Inc, Tucson, AZ). This process is automated and requires no RNA amplification or labeling, which might lead to more reproducible results. Gene-specific oligonucleotides are added to the sample and hybridize to the RNA present in solution. The oligonucleotides are added in excess to ensure that every molecule RNA capable of hybridizing does so. S1 nuclease is added to the hybridized sample buffer. The S1 nucleases will degrade any non-hybridized nucleic acid. This removes the non-hybridized portion of the targeted RNA, all of the non-targeted RNA and any excess oligonucleotides. The S1 nuclease is then inactivated. The RNA:DNA heteroduplexes are then

treated to remove the RNA portion of the duplex, leaving only the previously protected oligonucleotide probes. The resulting DNA oligos are essentially a library of the original RNA sample. Measurement of miRNAs is then performed using a qNPA ArrayPlate. The ArrayPlate spots are programmed using the appropriate gene-specific linkers which provide a specific capture sequence for the targeted miRNA. The qNPA protection oligos are added to the ArrayPlate wells and allowed to hybridize. The short, biotinylated miRNA protection oligos quantitatively hybridize to the programmed array spots. The array-bound biotinylated miRNA qNPA protection oligos are labeled with avidin-horseradish peroxidase (HRP) conjugate. The addition of a chemiluminescent substrate produces light, which is then quantitatively measured.

We performed whole qNPA transcriptome miRNA profiling on blood samples of varying ages: newborn (1 hour), mature adult (22 and 29 year olds) and elderly (73, 80 and 89 year olds). These age groups were initially selected since they represented significantly different stages of the aging cycling and also due to the availability of sample (the assay required microgram quantities, a potential significant limitation for future use with forensic samples). The miRNAs were successfully detected in the blood samples. The results from the whole transcriptome profiling allowed us to view entire miRNA populations at once rather than testing the individual primer assays as done with the miScript approach. Several candidates were identified that show possible age related expression patterns. Figures 9 and 10 show potential miRNA candidates for elderly (one candidate shown, miR-638) and newborns (two candidates shown, miR-197 (blue), miR-886-5p (yellow)), respectively. As can be seen from Figure 9, higher expression levels for miR-638 are observed for the three elderly samples (each run in duplicate with relatively consistent expression levels between duplicates). While the expression level of the miR-197 newborn candidates (blue, Figure 10) is higher than that of the miR-886-5p (yellow, Figure 10), both candidates show higher expression levels in the newborn samples compared to the other age groups. Additional candidates were identified for mature adult samples as well as candidates showing varying expression levels amongst the age groups (data not shown). For example, candidates were identified that showed high expression in all age groups except the oldest elderly sample. It is therefore possible that the expression of this miRNA decreases significantly after 80 years of age. Similarly, candidates were identified that were not expressed in newborns but were highly expressed in the 22-89 year old samples. Only a limited number of samples were included in the initial whole transcriptome profiling reactions and therefore it is possible that these candidates are expressed in mature adults and elderly only. While candidates showing these types of expression patterns are not the ideal presence/absence in a single age group, it is possible that they could be useful in combination with other age markers in order to add additional resolution between age groups.

The next stage of this analysis was to perform additional testing on the candidates identified from the whole transcriptome profiling (qSelect plate). This further analysis included the use of 28 bloodstains from donors of various ages (newborn - 0 and 1 hour, infant - 4 and 6 months, child - 4 years, adolescent - 12 and 14 years, adult – 17 to 41 years, middle-aged – 53 to 64 years, and elderly – 72 to 91 years). Each of these samples were run in triplicate. We included 14 potential age miRNA candidates for analysis with these samples based on those identified from the whole transcriptome data: miR-197, miR-886-5p, miR-590-3p, miR-1277, miR-638, miR-221, miR-423-3p, miR-768-3p, let-7a, miR-182, miR-15a, miR-223, miR-22, and miR-1297. The expression data from the whole transcriptome profiling was based on raw signal intensity values. No normalization strategy had been developed. While these samples were all from the same body fluid (blood), the expression data needed to be normalized in order to ensure

the accuracy of the expression data. Ideally, a miRNA expressed in equal levels across all age groups would be used for normalization purposes. Two potentially "housekeeping" miRNAs (miR548g and miR-1237) were identified in the whole transcriptome profiling data and therefore they were also included on the qSelect plate for analysis.

The 14 candidates selected for further analysis included potential candidates for the following: newborns (2 candidates), mature adults (2 candidates), elderly (1 candidate), other age "milestones" (remaining 9 candidates; high expression in all but newborns, high expression in all but elderly). After subsequent evaluation with the expanded set of age samples, the same expression trend as observed with the original whole transcriptome profiling was not observed for any of the candidates (using both raw and normalized data signals). The significant difference in expression observed between the whole transcriptome profiling and qSelect plate was somewhat surprising. In order to ensure that performed all possible analysis on the candidates identified from whole transcriptome profiling, we evaluated the expression of these same miRNAs with the Qiagen miScript assays in order to determine if they would show similar specificity for the targeted age groups. The specificity for the target age groups was not observed for any of the 14 candidates using the miScript assays. Based on these results, we did not conduct any further studies on qPNA assays. Due to the lack of consistency with the qPNA results, we chose to focus instead on the miScript miRNA assay work (see above sections).

4. Protein

Some of the most significant physiological changes in the human body occur during puberty and these are induced and regulated by several hormonal factors that are released by different organs and are then carried to the target tissues largely through the blood. This gives rise to short-term changes in the levels of these factors in the blood. The levels of these hormonal factors rise during this pubertal period and some of them remain elevated throughout a person's adult life while there are others that fall back to pre-pubertal levels and then remain low all through adulthood. Such changes are monitored by the medical community in order to test for any abnormalities in sexual development among adolescents. Based on this we considered the possibility of being able to observe such changes in levels of hormones within dried blood stains. If we were able to identify different levels of such proteins in our preserved blood samples of various ages, they might serve as potential biomarkers for the age of the sample donor.

We performed initial experiments using the IGFBP3 candidate (see earlier mRNA results section). We briefly tried a sandwich ELISA assay, which utilizes two sets of antibodies. Monoclonal capture antibodies are used to form a uniform coating on the bottom of the plate. These were commercially available antibodies that bind to a single specific epitope on the surface of the protein thus there is a one-one correlation between the number of antibodies and the number of protein molecules. Secondary antibodies then bind to the bound protein and a subsequent color reaction indicates the amount of protein present in the sample. A preliminary set of samples were tested and the results indicated that there was a distinct rise in the amount of candidate protein in the samples of age 14 years and over as compared to samples of a younger age (data not shown). However, after this preliminary study we proceeded to test a larger subset of our sample population using this assay. The clear distinction in protein levels that we had seen in our initial study was not consistently observed over a larger sample set (data not shown). Since protein analysis was not part of our original proposed study and would have required

extensive development and validation work, we did not perform any further protein studies and instead focused our efforts on the identification of mRNA and miRNA age biomarkers.

B. TABLES

Table 1. Representative example of RNA-seq data for obtained for 8 age samples

RefseqID	Description	FCHI	MCHI	FADO	MADO	FATD	MADT	FELD	MELD
		1 hr	33 mnths	12 yr	14 yr	31 yr	29 yr	71 yr	81 yr
NM_000016	Homo sapiens hemoglobin, gamma G (HBG2), mRNA	12261375	168901	291892	32243	29225	42554	189531	51543
NM_000017	Homo sapiens hemoglobin, gamma A (HBG1), mRNA	4641736	84892	539218	41360	69647	24277	189174	20658
NM_000018	Homo sapiens hemoglobin, alpha 1 (HBA1), mRNA	3299325	2362712	6483729	3768890	3919963	3774022	5229345	7294077
NM_000019	Homo sapiens hemoglobin, alpha 2 (HBA2), mRNA	1785611	1475780	2564215	2615816	1399430	3496556	1432549	4067793
NM_000021	Homo sapiens hemoglobin, beta (HBB), mRNA	1237792	3141802	3630807	5493442	5211962	6958520	1941070	3154641
NM_000025	Homo sapiens hemoglobin, delta (HBD), mRNA	725819	165162	93270	210249	73351	290162	63018	94623
NM_000026	Homo sapiens hemoglobin, epsilon 1 (HBE1), mRNA	496926	3846	6088	2100	1620	2324	5945	2134
NM_000027	Homo sapiens ribonuclease, RNase A family, 3 (RNASE3), mRNA	399561	1427	27218	564	4084	426	4066	218
NM_000030	Homo sapiens 28S ribosomal RNA (LOC100008589)	107283	370076	279175	289786	161665	253394	189639	185869
NM_000031	Homo sapiens 18S ribosomal RNA (LOC100008588)	99818	112910	182098	188253	105712	158903	45831	54272
NM_000032	Homo sapiens lysozyme (renal amyloidosis) (LYZ), mRNA	84407	149762	134368	151415	158641	93557	136378	10095
NM_000033	Homo sapiens tumor protein, translationally-controlled 1 (TPT1), mRNA	81402	150137	206859	249159	166103	123715	191718	26007
NM_000034	Homo sapiens G1 to S phase transition 1 (GSPT1), mRNA	65924	58193	35079	28219	31248	24302	44719	21433
NM_000035	Homo sapiens erythroid associated factor (ERAF), mRNA	54353	16838	19040	13844	35612	11229	111875	9380
NM_000036	Homo sapiens lipocalin 12 (LCN12), mRNA	37713	976	1502	199	180	374	773	565
NM_000037	Homo sapiens guanylate kinase 1 (GUK1), mRNA	31236	26864	25363	40366	50204	46627	31839	11224
NM_000038	Homo sapiens chromosome 20 open reading frame 54 (C20orf54), mRNA	29909	112	181	55	56	26	730	44
NM_000039	PREDICTED: Homo sapiens misc_RNA (LOC388707), miscRNA	27592	65735	51811	76690	68415	30790	50063	3401
NM_000045	PREDICTED: Homo sapiens misc_RNA (LOC728509), miscRNA	25563	30481	41183	47169	38914	21913	68798	1544
NM_000046	Homo sapiens adiponectin receptor 1 (ADIPOR1), mRNA	19822	21978	22287	24289	21152	21592	40183	36601
NM_000047	Homo sapiens paraoxonase 1 (PON1), mRNA	19678	145	64	23	27	26	68	38
NM_000051	Homo sapiens DEAD (Asp-Glu-Ala-Asp) box polypeptide 5 (DDX5), mRNA	19318	52132	50806	60005	82217	28243	28869	4330
NM_000052	Homo sapiens ribosomal protein S25 (RPS25), mRNA	19167	33471	49147	53339	35915	26527	26586	2018
NM_000053	Homo sapiens S100 calcium binding protein A9 (S100A9), mRNA	19022	24893	38209	33671	120510	9527	202851	53773
NM_000056	Homo sapiens eukaryotic translation initiation factor 1 (EIF1), mRNA	18995	95976	50189	72399	64774	21733	174842	21529
NM_000057	Homo sapiens ring finger protein 11 (RNF11), mRNA	18835	9119	4889	13358	6503	4126	16414	4911

Table 2. List of ~538 gene candidates evaluated as potential age biomarkers

ABAT	BGLAP	CHIA
ABL1	BIRC5	CIITA
ACD	BMAL1/ARNTL	c-KIT
ACTN3	C1QA	CLDN20
ACTA2	C1QB	CLEC1B
ADAM12	C18orf2	CLEC2
ADAMTS20	C19orf22	CLEC2A
ADIPOR1	C3	CLEC2B
AGGF1	C4A	CLU
AIF1	CA1	COL17A1
AJAP1	CABYR	COL1A1
AKT1	CAMK2D	COL3A1
AKT1S1	CAMP	COL5A1
ALOX15B-varA	CASP2	COL6A1
AMICA1	CAV1	COL6A2
AMID	CBL	COL4A5
AMPK/PRKAB1	CCL5	CORO1A
ANKH	CCM2	COX7A1
APEX1	CCND1	COX7A2
APOD	CCNE2	CR1
ARHGAP1	CCR4	CRP
ARMC7	CCR5	CSPG2/VCAN
ART3A	CCR7	CTBP1
ART3B	CCNE2	CST7
ASL	CD28	CTSA
ATF3	CD200	CTSB
ATN1	CD3D	CTSL
ATPAF2	CD3G	CTSK
ATOX1	CD3E	CTSS
ATR	CD55	CUX1
ATM	CD59	CYP17A1
AUF1	CD69	CYP19A1
AVEN	CDC6	CYP7B1
BACE1	CD86	CYTBC2
BAFF/TNFSF13B	CDC2	DCHS1
BAX	CDC6	DDB2
BAX-A/D	CDC25C	DEFA1
BAX-B	CDC42	DEFA3
BAX-D	CDKN1A	DHEA
BAX-E	CDKN1B	DKC1
BAX-S	CDKN2C	DLK1
BCKDHA	CDKN2A	DLX1
BCL2	CFIX	DMD-VAR1
BCL2A1	CGI-96	DMD-VAR2

DNCL2A01
DNCL2A-2
DNCL2A-3
DNPEP
DRD3
DYRK2
E2F1
E2IG2
ECGF1
EDARADD, VAR A
EDARADD, VAR B
EDF1
EEF1A1
EEF1A2
EFEMP1
EFHC1
EGFL7
EIF2KA1
EIF4E
ELAVL1
EMD
EPHX3
ERBP
ERCC1
ERCC8
EREG
ERF
ESR1
ESR2
ESRG
F2
FABP3
FAM83A
FAM100C
FBX024-V3
FCGR2B
FIL1
FKBP11
FLJ11078
FLJ20245
FLJ20421
FLJ22175
FLJ22672
FLJ35982
FLJ37440
FLJ38745
FLJ43159
FLJ35984
FLJ38628
FLJ39639
FLOT1
FLT1
FOXO1
FOXO3
FOXO4
GADD45A-V1
GADD45A-V3
GADD45B
GAL
GFI1
GFPT2
GGT1
GHR
GHRH
GLO1
GIMAP7
GLIPR1
GLIPR1L1
GNRH1
GNRH2
GNRHR
GPHA2
GPN1
GPR54/KISS1R
GPRIN3
GRIN1
GRIN2A
GRIN2B
GSTA1
GSTM3
GSTP1
GTF3C1
H17
H1FX
H2AX
H2AFX
HAGH
HBA1
HBA2
HBB
HBD
HBQ
HLA-DRB5
HIC2
HIF1A
HIST1HC1
HLA-DQB1
HLA-DRB5
HMOX1/HSP32\IGF1
HOMER3
HPCAL4
HRAS
HRG
HS3ST6
HSF1
HTATIP
HTR1E
HTR7
ICA1L
IFNG
IGF1R
IGFBP3
IGFBP5
IGJ
IGF1
IGF2
IL-2
IL-6
IL-7
IL-18
IL1A
IL1R2
INHA
INK4A/p16INK4A
INK4A/p14ARF
IRF1
ITGA2
ITGB1BP1
ITIH4
JAKMIP1
JARID1C
KCINN4
KDM5C
KHDC1
KIAA2022
KIR2DL4
KISS-1
KIT

KL
KLF13
KLF4
KRT6A
LAMTOR2
LAPTM5
LASS5
LATS1
LBXCOR1
LCK
LCP1
LEP
LHB
LHCGR
LIPA
LMNA
LOC283663
LOC152274
LOC284242
LOH11CR2
LONP1
LONP2
LRRC19
LYZ
MAD1L1
MAX
MCPH1
M-CSF
MDM2
MEPE
MET
MGC14288
MGC33894
MGC46336
MGC20460
MGC39650
MGST1, VAR 1B
MGST1, VAR 1C
MGST1, VAR 1D
MLL
MLL2
MMP-13
MMP-14
MMP-9
MRPL12
MTIF
MKRN1
MKNK2
MS4A3
MS4A4A
MTA1
MTX1
MYC
MYL9
NALP14
NBN
NCOR2
NDE1
NDUFA6
NDUFB11
NDUFS3
NEDD9
NFKBIA
NMI
NOS2A
NOTCH1
NOTCH2
NPPB
NPTX2
NRAS
NTS
OGG1
OPALIN
OPG
OPGL
ORAI3
OR56A1
OSCAR
OSGEP
OSM
OXTR
P2RY14
P53-A
P53-B
PADI4
PAQR6
PC-VAR2
PDCD1
PDCD10
PDCD11
PDCD1LG2
PDCD2L
PDCD4
PDCD5
PDCD6
PDCD6IP
PDCD7
PDE6D
PEMT
PEX5
PGC-1A
PGC-1B
PGR
PLD3
PLEKHA8
PIK3CA
PIK3R1
PIN1
PIP5K1A
PLA2G12A
PLCH1
PLXDC2
POLA1
POLA2
POLB
POLD1
POLE1
POLE2
POLE3
POLG
POLH
POLI
POLK
POLM
POLN
POLQ
POLR3F
POLR3K
POLS
POMC
POT1
PPARD
PPOX
PPP2R3A
PPRC1
PRAMEF4
PRDX5
PRDX6

PRKCA
PRL
PROP1
PRR5L
PRUNE2
PTGER4
PTH
PTMS
PTPN18
PTPRU
QDPR
RAD50
RAF1
RANKL
RAPA-2
RARA
RB1
RBBP8-V3
RBL1
RBL2
REA
RELA
RHOA
RHOC
RNH1
RPTOR
RSU1
RUNX1
RUNX2
RXRb
S100A9
S100P
S6K1
SA-b-gal
SCF
SELP
SEPX1
SEMA4A
SENP1
SESN1
SESN2
SESN3
SH3BGRL
SH3GL1
SHBG
SIRT1
SIRT2
SIRT3
SIRT4
SIRT5
SIRT6
SIRT7
SLC2A1
SLC39A4
SLC20A1
SLC25A39
SLC7A9
SMAP2
SMG5
SMG6
SMG7
SNCA
SNHG9
SNX9
SOX4
SPARC
SPATA1
SPINK1
SPINK5L3
SPINK-A
SPINK-B
SPP1
SPTRX-2
SRC
SRPX
SST
STK16
TACC1
TBC1
TCEAL2
TEKT2
TEP1
TERC
TERF2
TERT
TESC
TEX14
TFAP2BL1
TGFB1
TINF2
TMEM107
TNFAIP3
TNFRSF11A
TNFSF10
TNIP1
TNKS1BP1
TNRC18
TOB1
TOM1L1
TP53
TP53BP1
TP53BP2
TP53I3
TP73
TPST1
TPT1
TRAP
TRIB3
TRIM47
TRPC1
TSLL2
TTYH3
TUFT1
TXNDC5
UBA52
UBE2T
UPF1
UQCRFS1
UQCRQ
UNQ501
USP1
VAT1
VIM
VSIG2
WASF1
WISP2
WRN
XTP3TPA
ZER1
ZNF280C

Table 3. Expression Data for COL1A2 and IGFBP3

Classification	Age	IGFBP3 ΔCt	COL1A2 ΔCt
Newborn	1 hr	14.5	3.2
	1 hr	18.4	8.7
	1 day	19.0	6.0
	2 weeks	18.1	5.3
	17 day	22.2	7.2
	28 day	18.6	5.7
	63 days	15.6	2.9
	61 day	20.6	8.9
	17 month	17.8	12.1
Toddler	2 yr	15.4	4.0
	3 yr	17.5	7.3
	3 yr	16.1	7.4
	4 yr	18.5	10.3
	5 yr	14.9	12.4
Child	7 yr	12.0	22.1
	8 yr	3.9	16.0
	8 yr	7.4	18.0
	8 yr	16.5	6.0
	8 yr	9.1	21.8
	9 yr	7.5	15.7
	9 yr	7.0	16.5
	9 yr	14.8	4.9
	11 yr	5.8	18.2
	12 yr	5.4	16.6
Adolescent	13 yr	7.4	17.4
	13 yr	4.7	18.2
	13 yr	7.7	18.1
	15 yr	9.1	18.5
	17 yr	9.1	17.8
	15 yr	5.3	15.8
	16 yr	4.6	14.1
	16 yr	5.0	14.6
	17 yr	9.1	17.3

Classification	Age	IGFBP3 ΔCt	COL1A2 ΔCt
Adult	18 yr	6.5	18.7
	19 yr	9.6	18.6
	21 yr	5.1	20.7
	22 yr	6.5	10.2
	23 yr	6.9	16.4
	24 yr	4.7	16.9
	26 yr	8.0	19.1
	27 yr	5.9	19.2
	28 yr	7.0	16.7
	29 yr	7.0	17.5
	31 yr	9.7	17.3
	32 yr	5.2	13.7
	35 yr	5.3	14.8
	37 yr	2.6	16.2
	38 yr	5.5	16.4
	41 yr	4.8	19.1
	43 yr	7.5	17.1
Mature Adult	46 yr	6.8	18.8
	47 yr	6.0	20.7
	51 yr	9.2	21.3
	53 yr	7.8	16.2
	57 yr	6.7	15.8
	60 yr	6.7	20.0
Elderly	66 yr	6.2	20.6
	69 yr	3.4	16.3
	71 yr	5.7	15.3
	73 yr	4.6	14.0
	76 yr	5.2	17.8
	79 yr	3.3	14.3
	80 yr	2.3	16.3
	81 yr	4.9	17.4
	84 yr	7.8	16.8
	89 yr	8.7	19.4
	91 yr	6.0	15.5

Table 4. Primer sequences used in the developed mRNA profiling assays

Gene	Primer/Probe Sequence (5'-3')
IGFBP3	F: AGAACTTCTCCTCCGAGTCCAA
	R: CAGGTGATTCAGTGTGTCTTCCA
	P: NED-GGGAGACAGAATATGGTCCCTGCCG
COL1A2	F: GCATCCTTGGTTAGGGTCAATC
	R: CATGCCGTGACTTGAGACTCA
	P: VIC-AGTAACCACTGCTCCACTCTGGGTGG
B2M	F: GGCATTCCTGAAGCTGACA
	R: AAACCTGAATCTTTGGAGTACG
	P: VIC-TCGGGCCGAGATGTCTC
CCR7	F: CAGCCTTCCTGTGTGGTTTT
	R: TCCGTGACCTCATCTTGACA
RBBP8-v3	F: ACCCCCATGTCCGATACATA
	R: TTGGCCATTGGAGATTGAV
GFI1	F: TCCACACTGTCCACACACCT
	R: CTTGAAGCCTGTGTGTTTGC
GSTM3	F: TTCTGGGGAAATTCTCATGG
	R: CAGGCACTTGGGGTCAAATA
SENP1	F: GACTCCATGGGTGGGATAAA
	R: TGCTGAGGAATCTGGCTTTT
IGJ	F: GGAGTCCTGGCGGTTTTTAT
	R: AAGAACGGATGATCCTGGAA
UQCRFS1	F: AGCCTGTGTTGGACCTGAAG
	R: CGGCGGTATTCAGAGAAGTC
FCGR2B	F: GAGACCCTCCCTGAGAAACC
	R: ATTGTGTTCTCAGCCCCAAC
NDUFB11	F: CCCGAGGACGAAAACTTGTA
	R: CAAGGACCAGGATGATGGAG

F = forward primer
R = reverse primer
P = probe

Table 5. List of 84 potential miRNA age biomarkers identified by individual miRNA assay testing

Toddler
miR106b*
miR1204

Toddler-Adolescent
miR1290
miR3148
miR3149

Adolescent
miR30c
miR30e
miR302a
miR320a
miR588
miR200b*
miR3196
miR3689b*

Adult or Adult-Elderly
miR3921
miR513c
miR1284
miR621
miR345
miR337-3p
miR1251
miR581
miR598
miR1278
miR483-5p
miR302e
miR888
miR577
miR1827
miR627

Mature Adult
miR3120
miR517b
miR577c
miR941
miR337-5p
miR937
miR523
miR886-5p
miR1279
miR556-5p
miR1927
miR384
miR516a-5p
miR302f
miR1264
miR34b
miR584b
miR498
miR1224-3p

Mature Adult-Elderly
miR520a*
miR525

Elderly
miR520d-5p
miR33b_3
miR130a*
miR3162
miR3915
miR4275
miR4324
miR4327
miR4328
miR5483
miR513b
miR624
miR603
miR95
miR561
miR518b
miR518c
miR562
miR510
miR1207-3p
miR626
miR1265
miR1321
miR220c
miR612
miR657
miR802
miR455-3p
miR634
miR920
miR885-3p
miR512-5p
miR298
miR1324

Table 6. List of "A"-level Candidates Identified from miRNA Array Analysis

Plate-Position	Mature ID	Fold Change	Comments	Fold Regulation
3-A10	hsa-miR-363-3p	34.2677	A	34.2677
2-C01	hsa-miR-421	29.2475	A	29.2475
5-D03	hsa-miR-221-3p	20.5466	A	20.5466
9-D09	hsa-miR-320e	3.6512	A	3.6512
1-A06	hsa-let-7d-5p	2.9841	A	2.9841
1-D05	hsa-miR-19a-3p	2.3968	A	2.3968
1-D08	hsa-miR-423-5p	0.4426	A	-2.2592
3-F11	hsa-miR-130b-3p	0.3531	A	-2.8324
9-D07	hsa-miR-3607-3p	0.3071	A	-3.2565
2-E05	hsa-miR-324-5p	0.3016	A	-3.3159
4-F03	hsa-miR-151a-3p	0.2638	A	-3.7908
4-B12	hsa-miR-378a-3p	0.2245	A	-4.4553
5-E08	hsa-miR-324-3p	0.2012	A	-4.9695
11-C12	hsa-miR-4301	0.1941	A	-5.1524
11-C01	hsa-miR-4286	0.0888	A	-11.2593
4-C07	hsa-miR-629-5p	0.0763	A	-13.1099
8-D06	hsa-miR-1260a	0.0125	A	-80.1033
1-F06	hsa-miR-19b-3p	1.9663	A	1.9663
1-F12	hsa-miR-22-3p	1.7703	A	1.7703
1-G03	hsa-miR-29c-3p	1.5658	A	1.5658
9-D05	hsa-miR-2110	1.3605	A	1.3605
1-A09	hsa-miR-26a-5p	1.3549	A	1.3549
1-C07	hsa-miR-15a-5p	1.3274	A	1.3274
1-F02	hsa-let-7i-5p	1.0858	A	1.0858
1-C04	hsa-miR-223-3p	1.0184	A	1.0184
1-F08	hsa-miR-93-5p	0.8253	A	-1.2117
1-B07	hsa-miR-146a-5p	0.7968	A	-1.255
1-A05	hsa-miR-101-3p	0.788	A	-1.2691
1-A11	hsa-miR-26b-5p	0.7796	A	-1.2827
1-G08	hsa-miR-20a-5p	0.6287	A	-1.5907
1-F07	hsa-miR-17-5p	0.6172	A	-1.6202
10-D01	hsa-miR-3907	0.6166	A	-1.6218
10-D04	hsa-miR-378b	0.5921	A	-1.689
1-E07	hsa-miR-30a-5p	0.5842	A	-1.7119
1-E12	hsa-miR-30d-5p	0.55	A	-1.8183
1-G10	hsa-miR-7-5p	0.5373	A	-1.8613
1-F01	hsa-miR-191-5p	0.5291	A	-1.8901

C. FIGURES

Figure 1. 2D Scatterplot Assays for Potential Elderly mRNA Gene Candidates

2D scatterplots using normalized expression data (dCt – B2M): A) FCGR2B-C1QB, B) FCGR2B-C3, C) NDUFB11-FGCR2B, D) UQCRQ-UQCRFS1. For each assay, elderly samples are shown in purple and all other ages (newborn, toddler, child, adolescent, adult and mature adult) are shown in green.

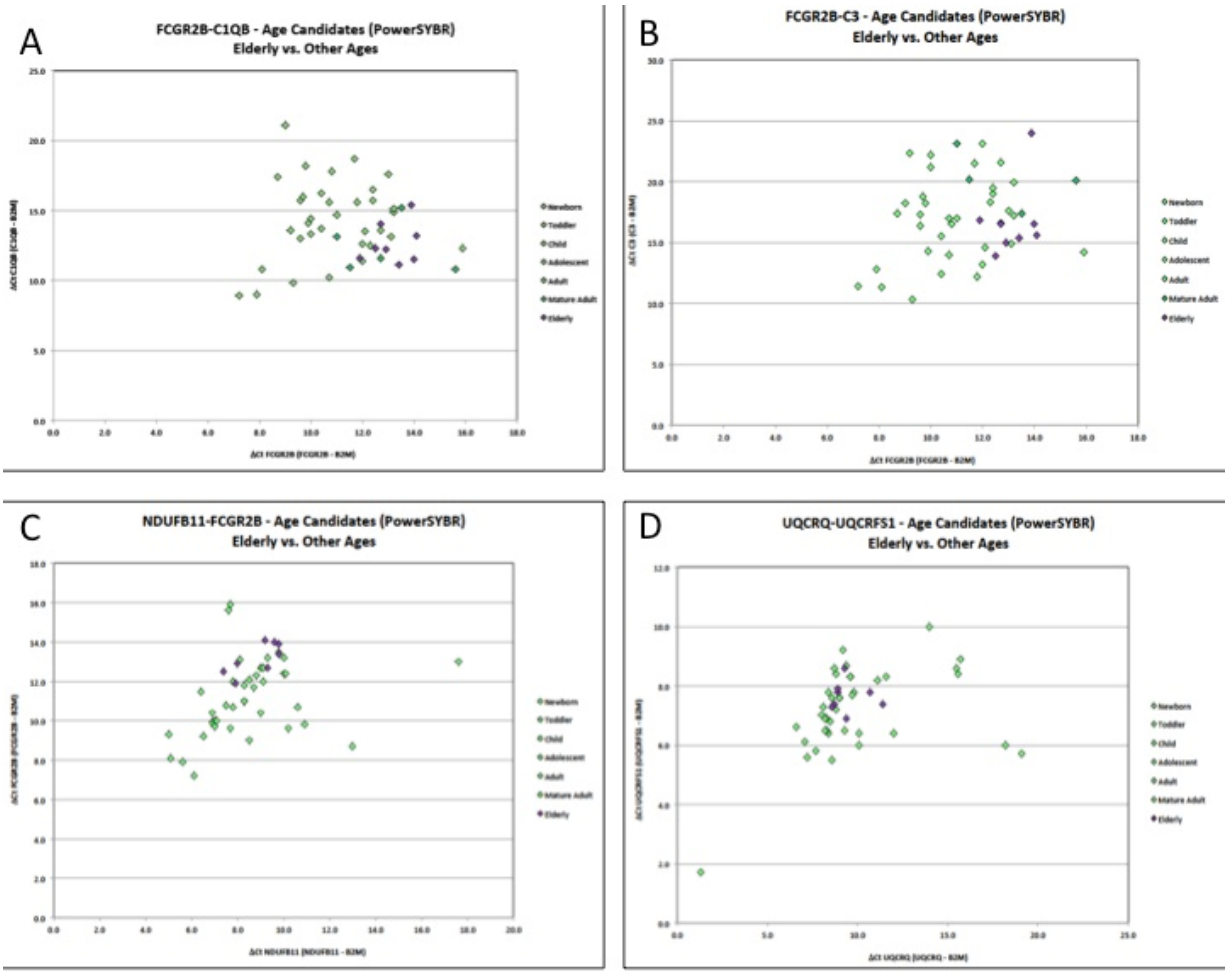


Figure 2. Potential mRNA Profiling Assay Workflow for the Determination of the Biological Age of a Bloodstain Donor

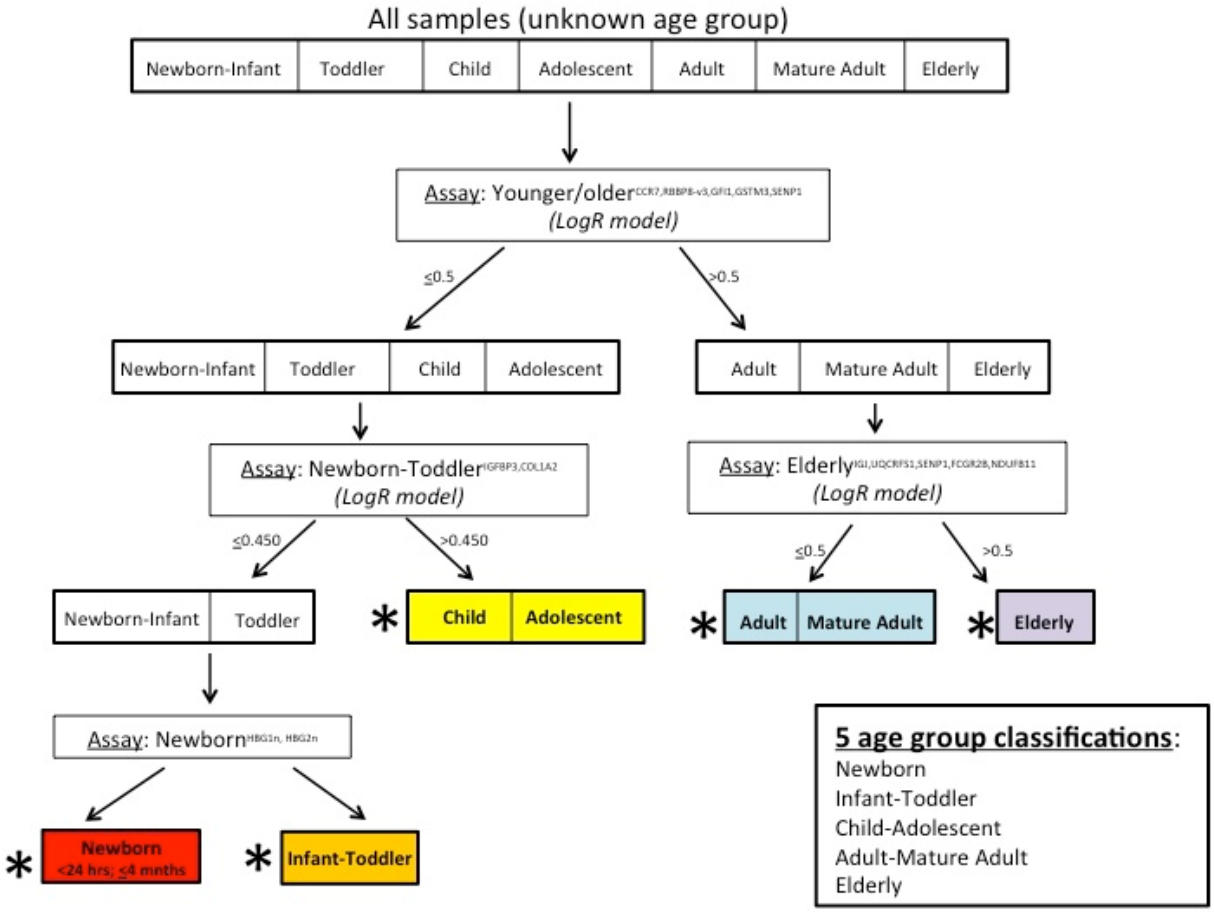


Figure 3. Alternative Potential mRNA Profiling Assay Workflow for the Determination of the Biological Age of a Bloodstain Donor

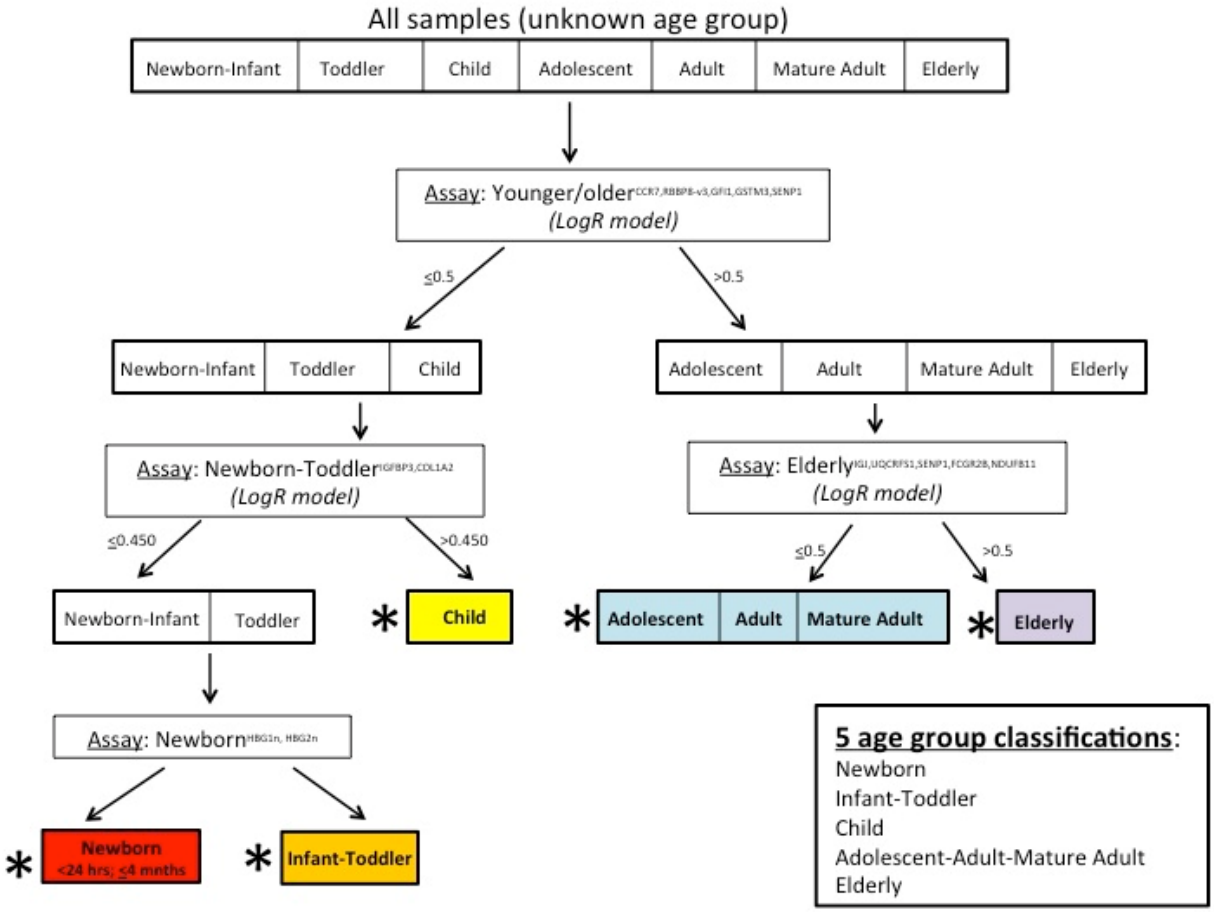


Figure 4. Evaluation of two potential miRNA candidates for age identification

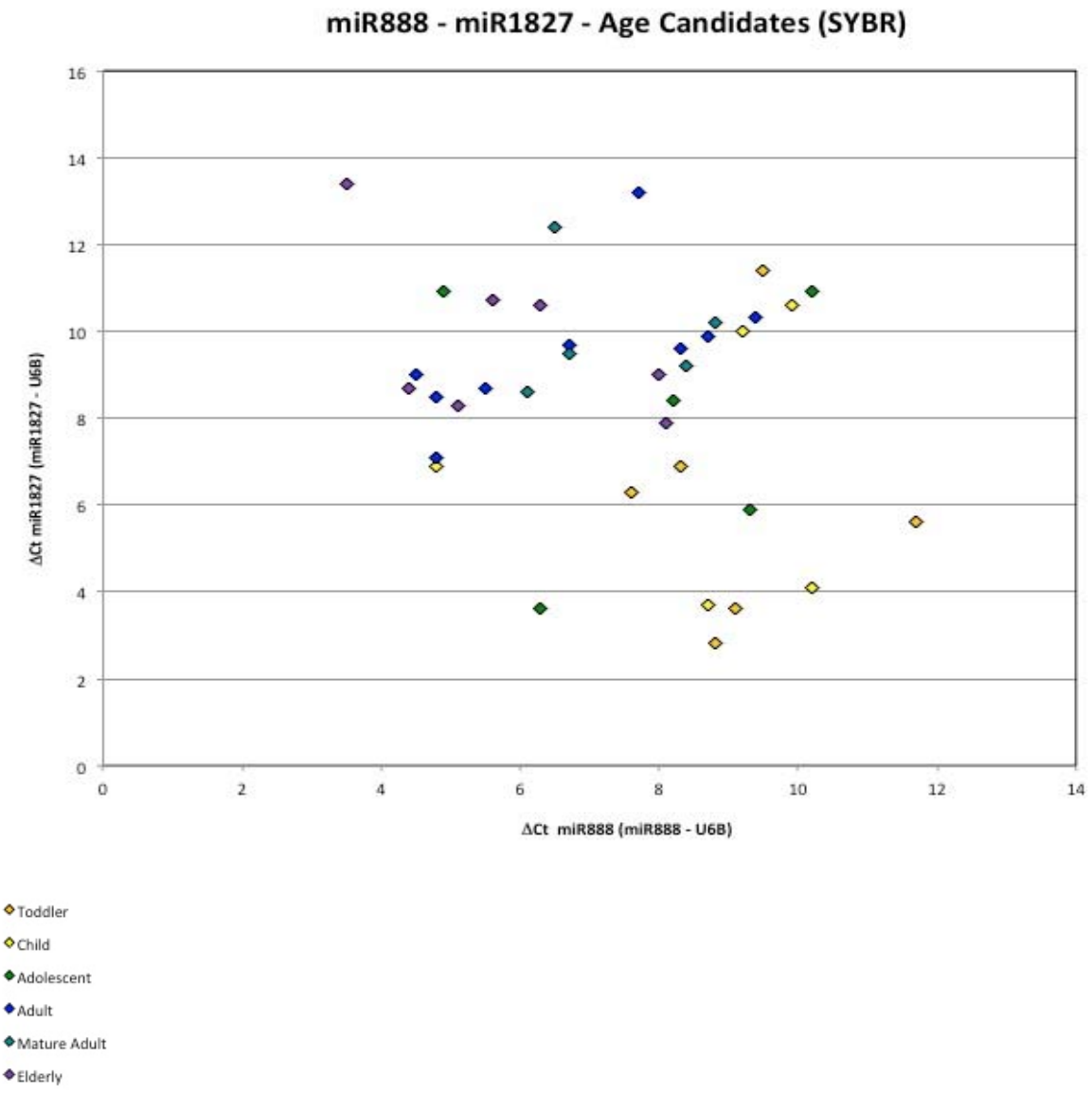


Figure 5. Data Summary from miRNA Array Analysis of an Adolescent Sample

C_t Range	Distribution of C_t Values	Average	ST DEV
	BD17-15y		
<25	31	31.00	N/A
25-30	92	92.00	N/A
30-35	207	207.00	N/A
Absent Calls	822	822.00	N/A
Percent Distribution of C_t Values			
<25	2.69%	2.69%	N/A
25-30	7.99%	7.99%	N/A
30-35	17.97%	17.97%	N/A
Absent Calls	71.35%	71.35%	N/A

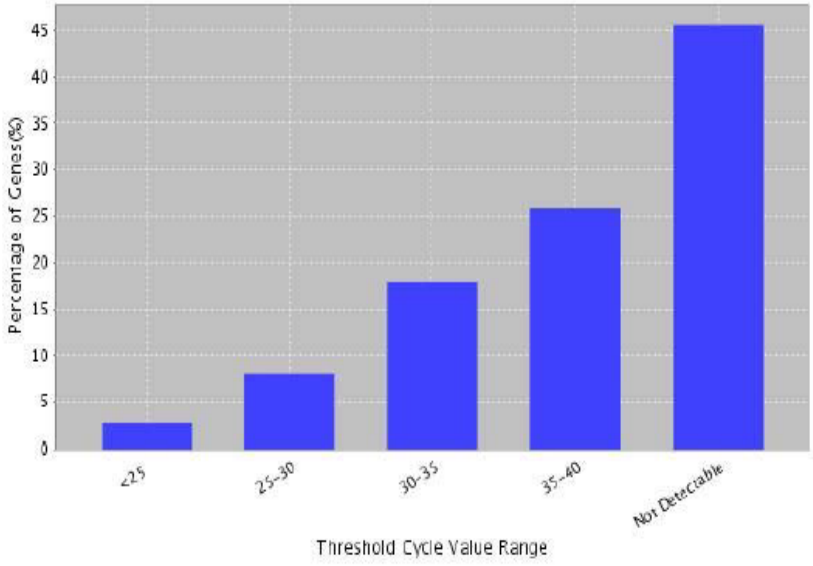


Figure 6. Data Summary from miRNA Array Analysis of an Elderly Sample

C_t Range	Distribution of C_t Values	Average	ST DEV
	BD64-89y		
<25	26	26.00	N/A
25-30	34	34.00	N/A
30-35	145	145.00	N/A
Absent Calls	947	947.00	N/A
Percent Distribution of C_t Values			
<25	2.26%	2.26%	N/A
25-30	2.95%	2.95%	N/A
30-35	12.59%	12.59%	N/A
Absent Calls	82.20%	82.20%	N/A

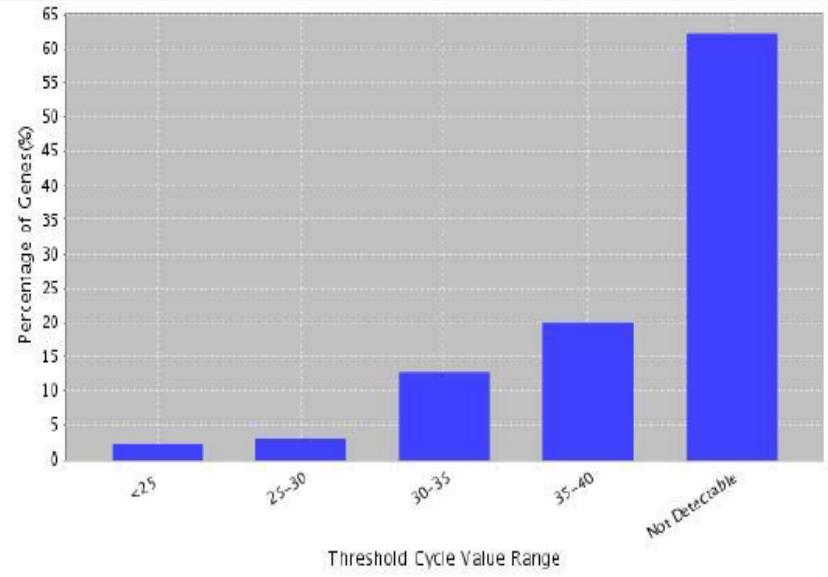
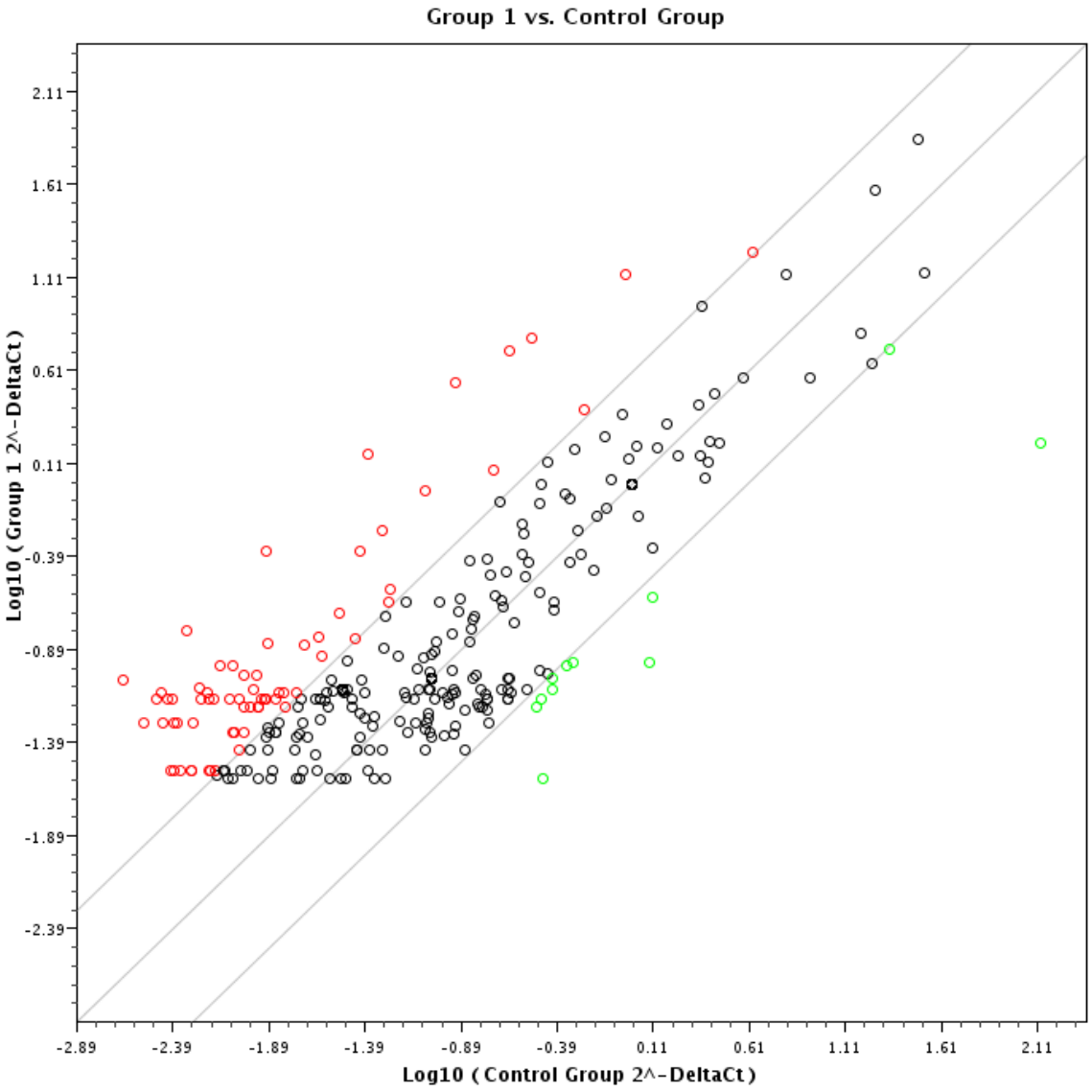


Figure 7. Scatterplot of expression fold changes for selected miRNA candidates from miRNA array analysis



Control group = adolescent; Group 1 = elderly

Figure 8. Evaluation of two potential miRNA candidates for age identification (identified from miRNA array analysis)

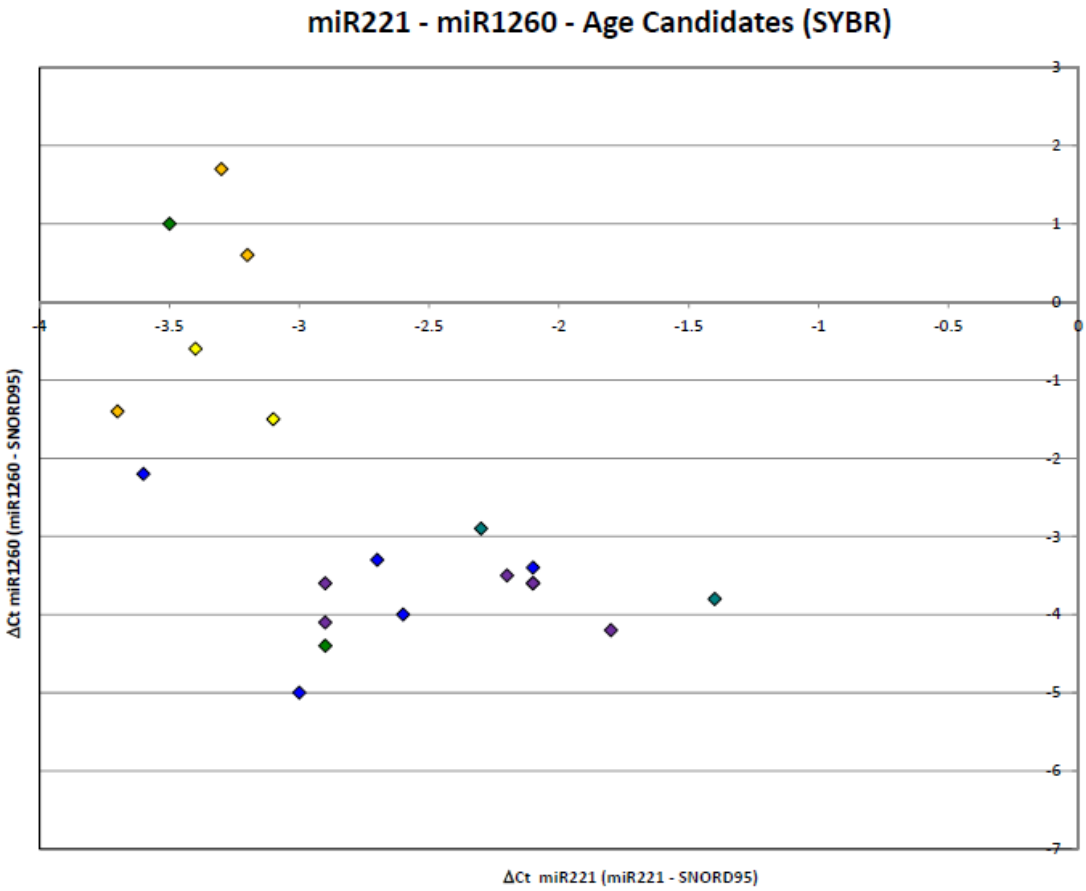


Figure 9. Elderly miRNA Candidate (miR-197) Identified Using qPNA Whole Transcriptome Profiling

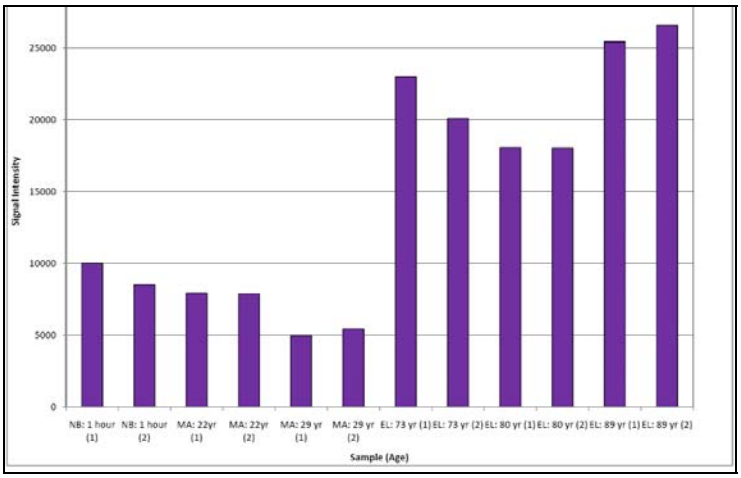
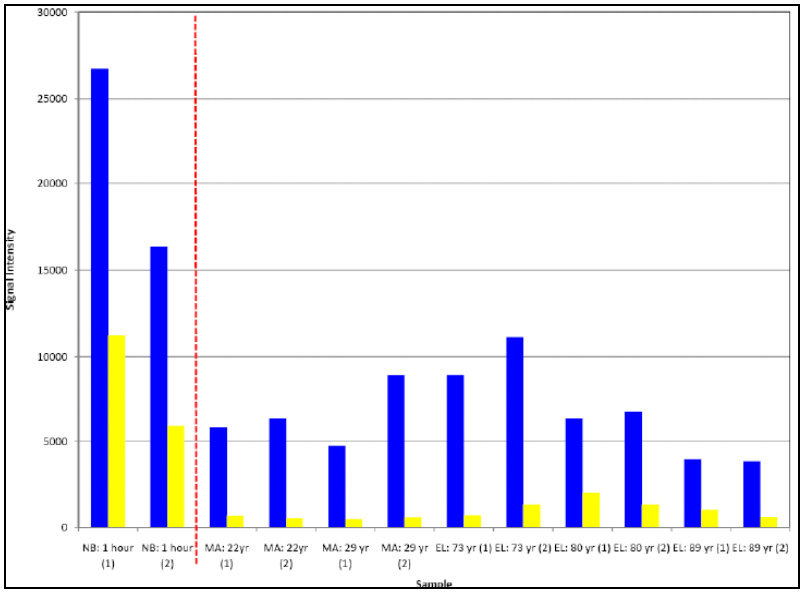


Figure 10. Newborn miRNA Candidates (miR-197, 886-5p) Identified Using qPNA Whole Transcriptome Profiling



miR-197 (blue); miR-886-5p (yellow)

IV. CONCLUSIONS

A. Discussion of findings

The ability to predict some key physical features of an individual by assaying appropriate specific biomarkers from crime scene samples would greatly aid criminal investigations. Recent progress has been made in identifying biomarkers associated with physical characteristics such as hair-, skin- and eye- pigmentation and bio-geographic ancestry but other physical features may also be similarly amenable to genetic analysis. An additional useful biometric that could provide important probative information is the biological age of an individual. The ability to determine whether a bloodstain donor is a newborn, infant, toddler, child, adolescent, adult or elderly individual could be useful in investigating certain cases such as kidnappings or for providing additional intelligence during terrorist investigations. However, there are no routinely used molecular assays for the determination of the biological age of a donor.

The human body ages over time and these changes are governed by a network of not so obvious, highly complex metabolic and genetic systems. These systems are activated or repressed during a person's lifetime and thus the metabolic state of these systems is expected to be representative of particular developmental ages. In previous work, we evaluated an approach to age determination that is based on developmental control of gene expression that occurs during these stages of human development. Through this initial work, we identified two novel gamma hemoglobin transcripts (HBG1n and HBG2n) that exhibit restricted expression in the blood of newborn children [115]. Individual qPCR assays were developed to measure both of these transcripts in forensic specimens. Adjustment of the primer concentrations in the qRT-PCR permitted the establishment of two temporally delimited assays, one of which was specific to blood from newborns age 4 months or under and the other of which was specific to blood from newborns who were just hours old (< 24 hours). The development of the newborn specific assay demonstrated the successful use of an RNA profiling approach to the determination of the biological age of an individual.

The aim of the current work was to further the evaluation of the use of RNA profiling methods for age determination by discovering markers for all major stages of life and developing a biomarker panel that could be used by the forensic community to predict the biological/chronological age of a bloodstain donor. This current work included a much more exhaustive and focused search and testing of genes from known biochemical pathways. Additionally, we uniquely utilized whole transcriptome profiling (RNA-Seq) analysis of samples at various stages in the human life cycle in order to obtain a more global perspective of gene expression and identify potential candidates that may not otherwise have been identified through other methods such as literature searches. Using both approaches to candidate identification, we identified and evaluated over 500 mRNA gene candidates as potential age biomarkers. The identification of age biomarkers in dried bloodstains proved to be extremely challenging. However, we have been able to successfully identify novel 9 mRNA candidates that when used in combination with four previously identified genes (HBG1n, HBG2n, IGFBP3 and COL1A2) permit a classification of bloodstains into five age classifications: 1) newborn, 2) infant-toddler, 3) child-adolescent, 4) adult-mature adult and 5) elderly (or alternatively 1) newborn, 2) infant-toddler, 3) child, 4) adolescent-adult-mature adult and 5) elderly, depending on further validation of one of the assays). This panel of ~15 biomarkers (including two housekeeping genes for normalization) is accomplished using four real time PCR assays, three of which are SYBR green

based and therefore do not require the use of more costly fluorescently labeled probes. Additionally, we have utilized a novel statistical approach for the analysis of the normalized expression data involving the use of logistic regression (LogR) models. The use of logistic regression has revolutionized analysis of expression data for not only this study, but also in other areas of our research such as miRNA profiling assays for the identification of forensically relevant body fluids [134]. The LogR models provide an objective "prediction" of the biological age of the bloodstain and permits the use of a multi-biomarker approach rather than only a two biomarker approach that is typically the case when using a real time PCR based platform. We had not performed such analysis when the initial newborn assay was developed and we are therefore hopeful that we can develop a suitable LogR model for this assay as well to further strengthen our ability to accurately predict biological age and have a definitive and uniform interpretation metric by which to make those predictions.

With the success of the mRNA profiling assays for age determination, we also sought to evaluate the use of miRNA profiling for the determination of the biological age of a bloodstain donor. In the current work, we have evaluated over 1,100 miRNAs to determine if any trends in regards to biological age can be observed. We evaluated these miRNAs using a small scale (i.e. small number of donors for initial testing) individual miRNA assay screening as well as more comprehensive miRNA array plates. As a result of this initial screening, we have identified 36 miRNAs as potential age biomarkers. We are in the process of further evaluating these miRNAs to determine if additional age assays can be developed. We are hopeful that the use of logistic regression analysis will permit a successful development of miRNA age profiling assays that could be used separately or in combination with the developed mRNA profiling assays for additional resolution of various age classifications.

The assays developed in the current work will all require additional validation before they could be used routinely for age determination. However, these results demonstrate the successful completion of all the major objectives of the current study: 1) evaluate of the use of RNA profiling methods for age determination by discovering markers for all major stages of life and 2) developing a biomarker panel that could be used by the forensic community to predict the biological/chronological age of a bloodstain donor. Further validation and refinement of these assays should result in a highly informative RNA profiling assay for the determination of the biological age of a bloodstain donor.

B. Implications for policy and practice

The ability to provide investigators with information as to the physical characteristics of an individual depositing a biological stain at a crime scene would provide important probative information to law enforcement investigators. The physical appearance of an individual is encoded in his/her genome as evidenced by the striking physical similarity of identical twins and so it should be possible to discern this information from a crime scene biological stain. The success of such studies would herald an exciting new post-genomic era in forensic biology in which novel phenotypic information about the donor of the body fluid could be obtained directly from the physiological stain itself by DNA and/or RNA analysis. This 'genetic eyewitness' evidence should assist law enforcement in developing or excluding suspects in criminal cases and could help clarify conventional eye witness accounts. Age is an important external visible characteristic that law enforcement investigators use during many criminal investigations to focus on a sub group of potential perpetrators. We are hopeful that an RNA profiling based age

identification assay will be developed with continued validation of the potential mRNA and miRNA candidates identified in this study. Even if not incorporated into stand-alone assays, perhaps the biomarkers identified in this study (or a sub-set thereof) could be incorporated into the massively parallel phenotypic traits assays envisioned as part of the next generation sequencing (NGS) forensic genomics revolution.

C. Implications for further research

In this work, we have successfully developed RNA profiling assays using 9 novel mRNA candidates, in addition to two housekeeping genes and four previously identified candidates (HBG1n, HBG2n, IGFBP3 and COL1A2), in order to permit a determination of which of several major biological age classifications a bloodstain door belongs to: newborn, toddler, child-adolescent, adult-mature adult and elderly. This was the main goal of the current study. We have employed logistic regression modeling that provides more definitive age classifications using the obtained expression data from this work. While the prototype assays have been developed, all assays and logR models will require additional validation work (testing of more samples, both known and unknown) before they would be ready for use in forensic casework. Since the newborn assay was developed prior to the use of our logistic regression approach to data interpretation, we would also like to develop a logR model for this assay as well. With successful development of this model, it may be possible to further the use of this type of statistical approach and use multinomial modeling in order to reduce the number of biomarkers needed for analysis.

In addition to the developed mRNA profiling assays, we have also completed an extensive evaluation of over 1,000 miRNAs as potential age biomarkers. We are still currently evaluating 36 potential miRNA candidates. Therefore we have only just begun to evaluate the use of miRNA age biomarkers and this will be an area of continued focus for us beyond the completion of the current study. Our previous work with miRNA profiling for body fluid identification has been quite successful and we are therefore hopeful that we will be able to develop suitable miRNA profiling assays for age determination as well. The use of logistic regression modeling has enhanced our ability to utilize RNA profiling assays and therefore we are confident that logistic regression analysis will have a significant impact on our success in developing potential miRNA profiling assays for age determination.

The original RNA-Seq data (mRNA) has provided a wealth of information for the age samples and has allowed us to identify significant numbers of potential candidates that may not have been identified through literature searches. However, we now have an in-house next generation sequence system available in our laboratory – the Ion Torrent PGM system. Since this technology differs from the original SOLID™ sequencing, we have begun preparations to perform additional whole transcriptome analysis of additional age samples. While the whole transcriptome sequencing provides valuable information for novel candidate identification, with the Ion Torrent PGM system, we have the opportunity to perform targeted sequencing reactions (i.e. AmpliSeq). Using this approach, multiplex primer pool of over 6,000 primer pairs can be designed to target a specified list of target genes. Therefore, it would be possibly to develop a targeted sequencing reaction using the various mRNA biomarkers identified in this study and continue to add to that network of biomarkers as new candidates are identified. It is very likely that an analysis of a complex network of genes will in fact be required for a determination of

biological age. While not performed in the current study, this will be an exciting future development towards a single assay for age determination.

V. REFERENCES

[1] V. Kastelic, B. Budowle, K. Drobnic, Validation of SRY marker for forensic casework analysis, J.Forensic Sci. 54 (2009) 551-555.

[2] A. Mannucci, K.M. Sullivan, P.L. Ivanov, P. Gill, Forensic application of a rapid and quantitative DNA sex test by amplification of the X-Y homologous gene amelogenin, Int.J.Legal Med. 106 (1994) 190-193.

[3] M. Steinlechner, B. Berger, H. Niederstatter, W. Parson, Rare failures in the amelogenin sex test, Int.J.Legal Med. 116 (2002) 117-120.

[4] H.E. Collins-Schramm, B. Chima, T. Morii, K. Wah, Y. Figueroa, L.A. Criswell, R.L. Hanson, W.C. Knowler, G. Silva, J.W. Belmont, M.F. Seldin, Mexican American ancestry-informative markers: examination of population structure and marker characteristics in European Americans, Mexican Americans, Amerindians and Asians, Hum.Genet. 114 (2004) 263-271.

[5] E. Giardina, I. Pietrangeli, C. Martinez-Labarga, C. Martone, A.F. de, A. Spinella, S.G. De, O. Rickards, G. Novelli, Haplotypes in SLC24A5 Gene as Ancestry Informative Markers in Different Populations, Curr.Genomics. 9 (2008) 110-114.

[6] M.F. Seldin, A.L. Price, Application of ancestry informative markers to association studies in European Americans, PLoS.Genet. 4 (2008) e5.

[7] C. Tian, D.A. Hinds, R. Shigeta, R. Kittles, D.G. Ballinger, M.F. Seldin, A genomewide single-nucleotide-polymorphism panel with high ancestry information for African American admixture mapping, Am.J.Hum.Genet. 79 (2006) 640-649.

[8] J.S. Allwood, S. Harbison, SNP model development for the prediction of eye colour in New Zealand, Forensic Sci.Int.Genet. 7 (2013) 444-452.

[9] W. Branicki, U. Brudnik, T. Kupiec, P. Wolanska-Nowak, A. Szczerbinska, A. Wojas-Pelc, Association of polymorphic sites in the OCA2 gene with eye colour using the tree scanning method, Ann.Hum.Genet. 72 (2008) 184-192.

[10] V. Kastelic, E. Pospiech, J. Draus-Barini, W. Branicki, K. Drobnic, Prediction of eye color in the Slovenian population using the IrisPlex SNPs, Croat.Med.J. 54 (2013) 381-386.

[11] M. Kayser, F. Liu, A.C. Janssens, F. Rivadeneira, O. Lao, D.K. van, M. Vermeulen, P. Arp, M.M. Jhamai, W.F. van Ijcken, J.T. den Dunnen, S. Heath, D. Zelenika, D.D. Despriet, C.C. Klaver, J.R. Vingerling, P.T. de Jong, A. Hofman, Y.S. Aulchenko, A.G. Uitterlinden, B.A. Oostra, C.M. van Duijn, Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene, Am.J.Hum.Genet. 82 (2008) 411-423.

[12] F. Liu, D.K. van, J.R. Vingerling, A. Hofman, A.G. Uitterlinden, A.C. Janssens, M. Kayser, Eye color and the prediction of complex phenotypes from genotypes, Curr.Biol. 19 (2009) R192-R193.

[13] J. Mengel-From, C. Borsting, J.J. Sanchez, H. Eiberg, N. Morling, Human eye colour and HERC2, OCA2 and MATP, Forensic Sci.Int.Genet. 4 (2010) 323-328.

[14] A. Pneuman, Z.M. Budimlija, T. Caragine, M. Prinz, E. Wurmbach, Verification of eye and skin color predictors in various populations, Leg.Med.(Tokyo). 14 (2012) 78-83.

[15] E. Pospiech, J. Draus-Barini, T. Kupiec, A. Wojas-Pelc, W. Branicki, Prediction of eye color from genetic data using Bayesian approach, J.Forensic Sci. 57 (2012) 880-886.
[16] O. Spichenok, Z.M. Budimlija, A.A. Mitchell, A. Jenny, L. Kovacevic, D. Marjanovic, T. Caragine, M. Prinz, E. Wurmbach, Prediction of eye and skin color in diverse populations using seven SNPs, Forensic Sci.Int.Genet. 5 (2011) 472-478.
[17] P. Sulem, D.F. Gudbjartsson, S.N. Stacey, A. Helgason, T. Rafnar, K.P. Magnusson, A. Manolescu, A. Karason, A. Palsson, G. Thorleifsson, M. Jakobsdottir, S. Steinberg, S. Palsson, F. Jonasson, B. Sigurgeirsson, K. Thorisdottir, R. Ragnarsson, K.R. Benediktsdottir, K.K. Aben, L.A. Kiemeney, J.H. Olafsson, J. Gulcher, A. Kong, U. Thorsteinsdottir, K. Stefansson, Genetic determinants of hair, eye and skin pigmentation in Europeans, Nat.Genet. 39 (2007) 1443-1452.
[18] S. Walsh, F. Liu, K.N. Ballantyne, O.M. van, O. Lao, M. Kayser, IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information, Forensic Sci.Int.Genet. 5 (2011) 170-180.
[19] S. Walsh, A. Lindenbergh, S.B. Zuniga, T. Sijen, K.P. de, M. Kayser, K.N. Ballantyne, Developmental validation of the IrisPlex system: determination of blue and brown iris colour for forensic intelligence, Forensic Sci.Int.Genet. 5 (2011) 464-471.
[20] S. Walsh, A. Wollstein, F. Liu, U. Chakravarthy, M. Rahu, J.H. Seland, G. Soubrane, L. Tomazzoli, F. Topouzis, J.R. Vingerling, J. Vioque, A.E. Fletcher, K.N. Ballantyne, M. Kayser, DNA-based eye colour prediction across Europe with the IrisPlex system, Forensic Sci.Int.Genet. 6 (2012) 330-340.
[21] S. Walsh, F. Liu, A. Wollstein, L. Kovatsi, A. Ralf, A. Kosiniak-Kamysz, W. Branicki, M. Kayser, The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA, Forensic Sci.Int.Genet. 7 (2013) 98-115.
[22] L. Yun, Y. Gu, H. Rajeevan, K.K. Kidd, Application of six IrisPlex SNPs and comparison of two eye color prediction systems in diverse Eurasia populations, Int.J.Legal Med.2014).
[23] W. Branicki, U. Brudnik, A. Wojas-Pelc, Interactions between HERC2, OCA2 and MC1R may influence human pigmentation phenotype, Ann.Hum.Genet. 73 (2009) 160-170.
[24] J. Han, P. Kraft, H. Nan, Q. Guo, C. Chen, A. Qureshi, S.E. Hankinson, F.B. Hu, D.L. Duffy, Z.Z. Zhao, N.G. Martin, G.W. Montgomery, N.K. Hayward, G. Thomas, R.N. Hoover, S. Chanock, D.J. Hunter, A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation, PLoS.Genet. 4 (2008) e1000074.
[25] K.L. Hart, S.L. Kimura, V. Mushailov, Z.M. Budimlija, M. Prinz, E. Wurmbach, Improved eye- and skin-color prediction based on 8 SNPs, Croat.Med.J. 54 (2013) 248-256.
[26] P.A. Kanetsky, J. Swoyer, S. Panossian, R. Holmes, D. Guerry, T.R. Rebbeck, A polymorphism in the agouti signaling protein gene is associated with human pigmentation, Am.J.Hum.Genet. 70 (2002) 770-775.
[27] R.L. Lamason, M.A. Mohideen, J.R. Mest, A.C. Wong, H.L. Norton, M.C. Aros, M.J. Jurynec, X. Mao, V.R. Humphreville, J.E. Humbert, S. Sinha, J.L. Moore, P. Jagadeeswaran, W. Zhao, G. Ning, I. Makalowska, P.M. McKeigue, D. O'donnell, R. Kittles, E.J. Parra, N.J. Mangini, D.J. Grunwald, M.D. Shriver, V.A. Canfield, K.C. Cheng, SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans, Science. 310 (2005) 1782-1786.

[28] O. Lao, J.M. de Gruijter, D.K. van, A. Navarro, M. Kayser, Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms, Ann.Hum.Genet. 71 (2007) 354-369.
[29] V. Neitzke-Montinelli, T.P. Urmenyi, E. Rondinelli, P.H. Cabello, R. Silva, R.S. Moura-Neto, Polymorphisms upstream of the melanocortin-1 receptor coding region are associated with human pigmentation variation in a Brazilian population, Am.J.Hum.Biol. 24 (2012) 853-855.
[30] P. Sulem, D.F. Gudbjartsson, S.N. Stacey, A. Helgason, T. Rafnar, M. Jakobsdottir, S. Steinberg, S.A. Gudjonsson, A. Palsson, G. Thorleifsson, S. Palsson, B. Sigurgeirsson, K. Thorisdottir, R. Ragnarsson, K.R. Benediktsdottir, K.K. Aben, S.H. Vermeulen, A.M. Goldstein, M.A. Tucker, L.A. Kiemeney, J.H. Olafsson, J. Gulcher, A. Kong, U. Thorsteinsdottir, K. Stefansson, Two newly identified genetic determinants of pigmentation in Europeans, Nat.Genet. 40 (2008) 835-837.
[31] G. Tully, Genotype versus phenotype: human pigmentation, Forensic Sci.Int.Genet. 1 (2007) 105-110.
[32] W. Branicki, U. Brudnik, J. Draus-Barini, T. Kupiec, A. Wojas-Pelc, Association of the SLC45A2 gene with physiological human hair colour variation, J.Hum.Genet. 53 (2008) 966-971.
[33] E.A. Grimes, P.J. Noake, L. Dixon, A. Urquhart, Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype, Forensic Sci.Int. 122 (2001) 124-129.
[34] M. Vaughn, O.R. van, S. Baindur-Hudson, Hair color measurement and variation, Am.J.Phys.Anthropol. 137 (2008) 91-96.
[35] B. Budowle, D.A. van, Forensically relevant SNP classes, Biotechniques. 44 (2008) 603-8, 610.
[36] D.F. Gudbjartsson, G.B. Walters, G. Thorleifsson, H. Stefansson, B.V. Halldorsson, P. Zusmanovich, P. Sulem, S. Thorlacius, A. Gylfason, S. Steinberg, A. Helgadottir, A. Ingason, V. Steinthorsdottir, E.J. Olafsdottir, G.H. Olafsdottir, T. Jonsson, K. Borch-Johnsen, T. Hansen, G. Andersen, T. Jorgensen, O. Pedersen, K.K. Aben, J.A. Witjes, D.W. Swinkels, H.M. den, B. Franke, A.L. Verbeek, D.M. Becker, L.R. Yanek, L.C. Becker, L. Tryggvadottir, T. Rafnar, J. Gulcher, L.A. Kiemeney, A. Kong, U. Thorsteinsdottir, K. Stefansson, Many sequence variants affecting diversity of adult human height, Nat.Genet. 40 (2008) 609-615.
[37] G. Lettre, A.U. Jackson, C. Gieger, F.R. Schumacher, S.I. Berndt, S. Sanna, S. Eyheramendy, B.F. Voight, J.L. Butler, C. Guiducci, T. Illig, R. Hackett, I.M. Heid, K.B. Jacobs, V. Lyssenko, M. Uda, M. Boehnke, S.J. Chanock, L.C. Groop, F.B. Hu, B. Isomaa, P. Kraft, L. Peltonen, V. Salomaa, D. Schlessinger, D.J. Hunter, R.B. Hayes, G.R. Abecasis, H.E. Wichmann, K.L. Mohlke, J.N. Hirschhorn, Identification of ten loci associated with height highlights new biological pathways in human growth, Nat.Genet. 40 (2008) 584-591.
[38] M. Mamada, T. Yorifuji, J. Yorifuji, K. Kurokawa, M. Kawai, T. Momoi, T. Nakahata, Fibrillin I gene polymorphism is associated with tall stature of normal individuals, Hum.Genet. 120 (2007) 733-735.
[39] K. Silventoinen, S. Sammalisto, M. Perola, D.I. Boomsma, B.K. Cornes, C. Davis, L. Dunkel, L.M. De, J.R. Harris, J.V. Hjelmborg, M. Luciano, N.G. Martin, J. Mortensen, L. Nistico, N.L. Pedersen, A. Skytthe, T.D. Spector, M.A. Stazi, G. Willemsen, J. Kaprio,

Heritability of adult body height: a comparative study of twin cohorts in eight countries, Twin.Res. 6 (2003) 399-408.
[40] P.M. Visscher, Sizing up human height variation, Nat.Genet. 40 (2008) 489-490.
[41] M.N. Weedon, G. Lettre, R.M. Freathy, C.M. Lindgren, B.F. Voight, J.R. Perry, K.S. Elliott, R. Hackett, C. Guiducci, B. Shields, E. Zeggini, H. Lango, V. Lyssenko, N.J. Timpson, N.P. Burtt, N.W. Rayner, R. Saxena, K. Ardlie, J.H. Tobias, A.R. Ness, S.M. Ring, C.N. Palmer, A.D. Morris, L. Peltonen, V. Salomaa, S.G. Davey, L.C. Groop, A.T. Hattersley, M.I. McCarthy, J.N. Hirschhorn, T.M. Frayling, A common variant of HMGA2 is associated with adult and childhood height in the general population, Nat.Genet. 39 (2007) 1245-1250.
[42] M.N. Weedon, H. Lango, C.M. Lindgren, C. Wallace, D.M. Evans, M. Mangino, R.M. Freathy, J.R. Perry, S. Stevens, A.S. Hall, N.J. Samani, B. Shields, I. Prokopenko, M. Farrall, A. Dominiczak, T. Johnson, S. Bergmann, J.S. Beckmann, P. Vollenweider, D.M. Waterworth, V. Mooser, C.N. Palmer, A.D. Morris, W.H. Ouwehand, J.H. Zhao, S. Li, R.J. Loos, I. Barroso, P. Deloukas, M.S. Sandhu, E. Wheeler, N. Soranzo, M. Inouye, N.J. Wareham, M. Caulfield, P.B. Munroe, A.T. Hattersley, M.I. McCarthy, T.M. Frayling, Genome-wide association analysis identifies 20 loci that influence adult height, Nat.Genet. 40 (2008) 575-583.
[43] T.L. Yang, D.H. Xiong, Y. Guo, R.R. Recker, H.W. Deng, Association analyses of CYP19 gene polymorphisms with height variation in a large sample of Caucasian nuclear families, Hum.Genet. 120 (2006) 119-125.
[44] J. Cui, J.L. Hopper, S.B. Harrap, Genes and family environment explain correlations between blood pressure and body mass index, Hypertension. 40 (2002) 7-12.
[45] V.L. Ellingrod, J.R. Bishop, J. Moline, Y.C. Lin, D.D. Miller, Leptin and leptin receptor gene polymorphisms and increases in body mass index (BMI) from olanzapine treatment in persons with schizophrenia, Psychopharmacol.Bull. 40 (2007) 57-62.
[46] T.M. Frayling, N.J. Timpson, M.N. Weedon, E. Zeggini, R.M. Freathy, C.M. Lindgren, J.R. Perry, K.S. Elliott, H. Lango, N.W. Rayner, B. Shields, L.W. Harries, J.C. Barrett, S. Ellard, C.J. Groves, B. Knight, A.M. Patch, A.R. Ness, S. Ebrahim, D.A. Lawlor, S.M. Ring, Y. Ben-Shlomo, M.R. Jarvelin, U. Sovio, A.J. Bennett, D. Melzer, L. Ferrucci, R.J. Loos, I. Barroso, N.J. Wareham, F. Karpe, K.R. Owen, L.R. Cardon, M. Walker, G.A. Hitman, C.N. Palmer, A.S. Doney, A.D. Morris, G.D. Smith, A.T. Hattersley, M.I. McCarthy, A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity, Science. 316 (2007) 889-894.
[47] V.S. Mattevi, V.M. Zembrzuski, M.H. Hutz, A resistin gene polymorphism is associated with body mass index in women, Hum.Genet. 115 (2004) 208-212.
[48] J.M. McCaffery, G.D. Papandonatos, D.S. Bond, M.J. Lyons, R.R. Wing, Gene X environment interaction of vigorous exercise and body mass index among male Vietnam-era twins, Am.J.Clin.Nutr. 89 (2009) 1011-1018.
[49] S. Alappat, Z.Y. Zhang, Y.P. Chen, Msx homeobox gene family and craniofacial development, Cell Res. 13 (2003) 429-442.
[50] T. Frudakis, Molecular Photofitting, Academic Press, 2007.
[51] P. Kraus, T. Lufkin, Dlx homeobox gene control of mammalian limb and craniofacial development, Am.J.Med.Genet.A. 140 (2006) 1366-1374.
[52] D.M. Noden, Vertebrate craniofacial development: novel approaches and new dilemmas, Curr.Opin.Genet.Dev. 2 (1992) 576-581.

[53] M. Tassabehji, P. Hammond, A. Karmiloff-Smith, P. Thompson, S.S. Thorgeirsson, M.E. Durkin, N.C. Popescu, T. Hutton, K. Metcalfe, A. Rucka, H. Stewart, A.P. Read, M. Maconochie, D. Donnai, GTF2IRD1 in craniofacial development of humans and mice, Science. 310 (2005) 1184-1187.
[54] R. Arking, NetLibrary Inc, The biology of aging: observations and principles, 2006.
[55] L. Guarente, Molecular Biology of Aging, Cold Springs Harbour, 2008.
[56] A. Macierira-Coelho, Molecular basis of aging, Boca Raton, FL, 1995.
[57] L. Guarente, C. Kenyon, Genetic pathways that regulate ageing in model organisms, Nature. 408 (2000) 255-262.
[58] A. Weismann Life and Death. in: A.E.S.Edward Bagnall Poulton (Ed.) Essays Upon Heredity and Kindred Biological Problems, Clarendon Press, Oxford, 1891, pp. 111-157.
[59] A. Cerami, Hypothesis. Glucose as a mediator of aging, J.Am.Geriatr.Soc. 33 (1985) 626-634.
[60] A. Cerami, H. Vlassara, M. Brownlee, Glucose and aging, Sci.Am. 256 (1987) 90-96.
[61] D.R. Sell Long Lived Proteins: Extracellular Matrix (collagen, elastins, proteoglycans) and lens crystallins). in: E. J. Masoro (Ed.) Handbook of Physiology. Section 11: aging, Oxford Press, New York, 1995, pp. 235-308.
[62] G.M. Martin, M. Fry, L.A. Loeb Somatic Mutation and Aging in Mammalian Cells. Molecular Biology of Aging: Gene Stability and Gene Expression, Raven Press, New York, 1985, pp. 7-21.
[63] L. Szilard, On the nature of the aging process, Proc.Natl.Acad.Sci.U.S.A. 45 (1959) 30-45.
[64] S. Agarwal, R.S. Sohal, Relationship between susceptibility to protein oxidation, aging, and maximum life span potential of different species, Exp.Gerontol. 31 (1996) 365-372.
[65] D.E. Promislow, DNA repair and the evolution of longevity: a critical analysis, J.Theor.Biol. 170 (1994) 291-300.
[66] A. Szakmary, S.M. Huang, D.T. Chang, P.A. Beachy, M. Sander, Overexpression of a Rrp1 transgene reduces the somatic mutation and recombination frequency induced by oxidative DNA damage in Drosophila melanogaster, Proc.Natl.Acad.Sci.U.S.A. 93 (1996) 1607-1612.
[67] D.C. Wallace, Mitochondrial diseases in man and mouse, Science. 283 (1999) 1482-1488.
[68] V. Lundblad, J.W. Szostak, A mutant with a defect in telomere elongation leads to senescence in yeast, Cell. 57 (1989) 633-643.
[69] E. Ponsot, J. Lexell, F. Kadi, Skeletal muscle telomere length is not impaired in healthy physically active old women and men, Muscle Nerve. 37 (2008) 467-472.
[70] D. Zekry, F.R. Herrmann, I. Irminger-Finger, L. Ortolan, C. Genet, A.M. Vitale, J.P. Michel, G. Gold, K.H. Krause, Telomere length is not predictive of dementia or MCI conversion in the oldest old, Neurobiol.Aging. 31 (2010) 719-720.
[71] J.L. Jameson, Rites of passage through puberty: a complex genetic ensemble, Proc.Natl.Acad.Sci.U.S.A. 104 (2007) 17247-17248.
[72] M.R. Palmert, P.A. Boepple, Variation in the timing of puberty: clinical spectrum and genetic investigation, J.Clin.Endocrinol.Metab. 86 (2001) 2364-2368.
[73] N.D. Quinton, R.F. Smith, P.E. Clayton, M.S. Gill, S. Shalet, S.K. Justice, S.A. Simon, S. Walters, M.C. Postel-Vinay, A.I. Blakemore, R.J. Ross, Leptin binding activity changes

with age: the link between leptin and puberty, J.Clin.Endocrinol.Metab. 84 (1999) 2336-2341.
[74] J. Domen, I.L. Weissman, Hematopoietic stem cells need two signals to prevent apoptosis; BCL-2 can provide one of these, Kitl/c-Kit signaling the other, J.Exp.Med. 192 (2000) 1707-1718.
[75] C. Levy, M. Khaled, D.E. Fisher, MITF: master regulator of melanocyte development and melanoma oncogene, Trends Mol.Med. 12 (2006) 406-414.
[76] G.G. McGill, M. Horstmann, H.R. Widlund, J. Du, G. Motyckova, E.K. Nishimura, Y.L. Lin, S. Ramaswamy, W. Avery, H.F. Ding, S.A. Jordan, I.J. Jackson, S.J. Korsmeyer, T.R. Golub, D.E. Fisher, Bcl2 regulation by the melanocyte master regulator Mitf modulates lineage survival and melanoma cell viability, Cell. 109 (2002) 707-718.
[77] V. Humphreys, R.M. Martin, B. Ratcliffe, S. Duthie, S. Wood, D. Gunnell, A.R. Collins, Age-related increases in DNA repair and antioxidant protection: a comparison of the Boyd Orr Cohort of elderly subjects with a younger population sample, Age Ageing. 36 (2007) 521-526.
[78] E.C. Berglund, A. Kiialainen, A.C. Syvanen, Next-generation sequencing technologies and applications for human genetic history and forensics, Investig.Genet. 2 (2011) 23.
[79] E.R. Mardis, Next-generation DNA sequencing methods, Annu.Rev.Genomics Hum.Genet. 9 (2008) 387-402.
[80] J.M. Rothberg, W. Hinz, T.M. Rearick, J. Schultz, W. Mileski, M. Davey, J.H. Leamon, K. Johnson, M.J. Milgrew, M. Edwards, J. Hoon, J.F. Simons, D. Marran, J.W. Myers, J.F. Davidson, A. Branting, J.R. Nobile, B.P. Puc, D. Light, T.A. Clark, M. Huber, J.T. Branciforte, I.B. Stoner, S.E. Cawley, M. Lyons, Y. Fu, N. Homer, M. Sedova, X. Miao, B. Reed, J. Sabina, E. Feierstein, M. Schorn, M. Alanjary, E. Dimalanta, D. Dressman, R. Kasinskas, T. Sokolsky, J.A. Fidanza, E. Namsaraev, K.J. McKernan, A. Williams, G.T. Roth, J. Bustillo, An integrated semiconductor device enabling non-optical genome sequencing, Nature. 475 (2011) 348-352.
[81] J. Shendure, H. Ji, Next-generation DNA sequencing, Nat.Biotechnol. 26 (2008) 1135-1145.
[82] B.T. Wilhelm, S. Marguerat, S. Watt, F. Schubert, V. Wood, I. Goodhead, C.J. Penkett, J. Rogers, J. Bahler, Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution, Nature. 453 (2008) 1239-1243.
[83] B. Wold, R.M. Myers, Sequence census methods for functional genomics, Nat.Methods. 5 (2008) 19-21.
[84] S.A. McCarroll, C.T. Murphy, S. Zou, S.D. Pletcher, C.S. Chin, Y.N. Jan, C. Kenyon, C.I. Bargmann, H. Li, Comparing genomic expression patterns across species identifies shared transcriptional profile in aging, Nat.Genet. 36 (2004) 197-204.
[85] J.M. Zahn, R. Sonu, H. Vogel, E. Crane, K. Mazan-Mamczarz, R. Rabkin, R.W. Davis, K.G. Becker, A.B. Owen, S.K. Kim, Transcriptional profiling of aging in human muscle reveals a common aging signature, PLoS.Genet. 2 (2006) e115.
[86] S.K. Kim, Common aging pathways in worms, flies, mice and humans, J.Exp.Biol. 210 (2007) 1607-1612.
[87] S. Someya, T. Yamasoba, T.A. Prolla, M. Tanokura, Genes encoding mitochondrial respiratory chain components are profoundly down-regulated with aging in the cochlea of DBA/2J mice, Brain Res. 1182 (2007) 26-33.

[88] J.M. Zahn, S. Poosala, A.B. Owen, D.K. Ingram, A. Lustig, A. Carter, A.T. Weeraratna, D.D. Taub, M. Gorospe, K. Mazan-Mamczarz, E.G. Lakatta, K.R. Boheler, X. Xu, M.P. Mattson, G. Falco, M.S. Ko, D. Schlessinger, J. Firman, S.K. Kummerfeld, W.H. Wood, III, A.B. Zonderman, S.K. Kim, K.G. Becker, AGEMAP: a gene expression database for aging in mice, PLoS.Genet. 3 (2007) e201.
[89] J.M. Zahn, S.K. Kim, Systems biology of aging in four species, Curr.Opin.Biotechnol. 18 (2007) 355-359.
[90] E.M. Blalock, J.W. Geddes, K.C. Chen, N.M. Porter, W.R. Markesbery, P.W. Landfield, Incipient Alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses, Proc.Natl.Acad.Sci.U.S.A. 101 (2004) 2173-2178.
[91] H.B. Fraser, P. Khaitovich, J.B. Plotkin, S. Paabo, M.B. Eisen, Aging and gene expression in the primate brain, PLoS.Biol. 3 (2005) e274.
[92] W.S. Liang, T. Dunckley, T.G. Beach, A. Grover, D. Mastroeni, D.G. Walker, R.J. Caselli, W.A. Kukull, D. McKeel, J.C. Morris, C. Hulette, D. Schmechel, G.E. Alexander, E.M. Reiman, J. Rogers, D.A. Stephan, Gene expression profiles in anatomically and functionally distinct regions of the normal aged human brain, Physiol Genomics. 28 (2007) 311-322.
[93] T. Lu, Y. Pan, S.Y. Kao, C. Li, I. Kohane, J. Chan, B.A. Yankner, Gene regulation and DNA damage in the ageing human brain, Nature. 429 (2004) 883-891.
[94] R. Ricciarelli, C. d'Abramo, S. Massone, U. Marinari, M. Pronzato, M. Tabaton, Microarray analysis in Alzheimer's disease and normal aging, IUBMB.Life. 56 (2004) 349-354.
[95] A. Melk, E.S. Mansfield, S.C. Hsieh, T. Hernandez-Boussard, P. Grimm, D.C. Rayner, P.F. Halloran, M.M. Sarwal, Transcriptional analysis of the molecular basis of human kidney aging using cDNA microarray profiling, Kidney Int. 68 (2005) 2667-2679.
[96] G.E. Rodwell, R. Sonu, J.M. Zahn, J. Lund, J. Wilhelmy, L. Wang, W. Xiao, M. Mindrinos, E. Crane, E. Segal, B.D. Myers, J.D. Brooks, R.W. Davis, J. Higgins, A.B. Owen, S.K. Kim, A transcriptional profile of aging in the human kidney, PLoS.Biol. 2 (2004) e427.
[97] P.G. Giresi, E.J. Stevenson, J. Theilhaber, A. Koncarevic, J. Parkington, R.A. Fielding, S.C. Kandarian, Identification of a molecular signature of sarcopenia, Physiol Genomics. 21 (2005) 253-263.
[98] S. Welle, A. Brooks, C.A. Thornton, Senescence-related changes in gene expression in muscle: similarities and differences between mice and men, Physiol Genomics. 5 (2001) 67-73.
[99] S. Welle, A.I. Brooks, J.M. Delehanty, N. Needler, C.A. Thornton, Gene expression profile of aging in human muscle, Physiol Genomics. 14 (2003) 149-159.
[100] S. Welle, A.I. Brooks, J.M. Delehanty, N. Needler, K. Bhatt, B. Shah, C.A. Thornton, Skeletal muscle gene expression profiles in 20-29 year old and 65-71 year old women, Exp.Gerontol. 39 (2004) 369-377.
[101] R.P. Thomas, M. Guigneaux, T. Wood, B.M. Evers, Age-associated changes in gene expression patterns in the liver, J.Gastrointest.Surg. 6 (2002) 445-453.
[102] J.R. Hawse, J.F. Hejtmancik, J. Horwitz, M. Kantorow, Identification and functional clustering of global gene expression differences between age-related cataract and clear human lenses and aged human lenses, Exp.Eye Res. 79 (2004) 935-940.

[103] F. Segev, O. Mor, A. Segev, M. Belkin, E.I. Assia, Downregulation of gene expression in the ageing lens: a possible contributory factor in senile cataract, Eye (Lond). 19 (2005) 80-85.
[104] S. Yoshida, B.M. Yashar, S. Hiriyanna, A. Swaroop, Microarray analysis of gene expression in the aging human retina, Invest Ophthalmol.Vis.Sci. 43 (2002) 2554-2560.
[105] T. Lener, P.R. Moll, M. Rinnerthaler, J. Bauer, F. Aberger, K. Richter, Expression profiling of aging in the human skin, Exp.Gerontol. 41 (2006) 387-397.
[106] Q. Tan, K. Christensen, L. Christiansen, H. Frederiksen, L. Bathum, J. Dahlgaard, T.A. Kruse, Genetic dissection of gene expression observed in whole blood samples of elderly Danish twins, Hum.Genet. 117 (2005) 267-274.
[107] R.D. Visala, G.M. Boyle, P.G. Parsons, K. Watson, G.L. Jones, Influence of ageing, heat shock treatment and in vivo total antioxidant status on gene-expression profile and protein synthesis in human peripheral lymphocytes, Mech.Ageing Dev. 124 (2003) 55-69.
[108] D. Bryder, D.J. Rossi, I.L. Weissman, Hematopoietic stem cells: the paradigmatic tissue-specific stem cell, Am.J.Pathol. 169 (2006) 338-346.
[109] C. Beghe, A. Wilson, W.B. Ershler, Prevalence and outcomes of anemia in geriatrics: a systematic review of the literature, Am.J.Med. 116 Suppl 7A (2004) 3S-10S.
[110] M.A. Lichtman, J.M. Rowe, The relationship of patient age to the pathobiology of the clonal myeloid diseases, Semin.Oncol. 31 (2004) 185-197.
[111] P.J. Linton, K. Dorshkind, Age-related changes in lymphocyte development and function, Nat.Immunol. 5 (2004) 133-139.
[112] A. Nijnik, L. Woodbine, C. Marchetti, S. Dawson, T. Lambe, C. Liu, N.P. Rodrigues, T.L. Crockford, E. Cabuy, A. Vindigni, T. Enver, J.I. Bell, P. Slijepcevic, C.C. Goodnow, P.A. Jeggo, R.J. Cornall, DNA repair is limiting for haematopoietic stem cells during ageing, Nature. 447 (2007) 686-690.
[113] D.J. Rossi, D. Bryder, J. Seita, A. Nussenzweig, J. Hoeijmakers, I.L. Weissman, Deficiencies in DNA damage repair limit the function of haematopoietic stem cells with age, Nature. 447 (2007) 725-729.
[114] S.M. Chambers, C.A. Shaw, C. Gatza, C.J. Fisk, L.A. Donehower, M.A. Goodell, Aging hematopoietic stem cells decline in function and exhibit epigenetic dysregulation, PLoS.Biol. 5 (2007) e201.
[115] M. Alvarez, J. Ballantyne, The identification of newborns using messenger RNA profiling analysis, Anal.Biochem. 357 (2006) 21-34.
[116] M. Alvarez, J. Ballantyne, Identification of four novel developmentally regulated gamma hemoglobin mRNA isoforms, Exp.Hematol. 37 (2009) 285-293.
[117] K. Kerschan-Schindl, G. Hawa, S. Kudlacek, W. Woloszczuk, P. Pietschmann, Serum levels of cathepsin K decrease with age in both women and men, Exp.Gerontol. 40 (2005) 532-535.
[118] P. Scaffidi, T. Misteli, Lamin A-dependent nuclear defects in human aging, Science. 312 (2006) 1059-1063.
[119] W. Wang, X. Yang, V.J. Cristofalo, N.J. Holbrook, M. Gorospe, Loss of HuR is linked to reduced expression of proliferative genes during replicative senescence, Mol.Cell Biol. 21 (2001) 5889-5898.

[120] K. Shiga, H. Yamamoto, H. Okamoto, Isolation and characterization of the human homologue of rig and its pseudogenes: the functional gene has features characteristic of housekeeping genes, Proc.Natl.Acad.Sci.U.S.A. 87 (1990) 3594-3598.
[121] P.T. Haak, J.V. Busik, E.J. Kort, M. Tikhonenko, N. Paneth, J.H. Resau, Archived unfrozen neonatal blood spots are amenable to quantitative gene expression analysis, Neonatology. 95 (2009) 210-216.
[122] E.K. Hanson, H. Lubenow, J. Ballantyne, Identification of forensically relevant body fluids using a panel of differentially expressed microRNAs, Anal.Biochem. 387 (2009) 303-314.
[123] E.K. Hanson, J. Ballantyne, Circulating microRNA for the identification of forensically relevant body fluids, Methods Mol.Biol. 1024 (2013) 221-234.
[124] J. Juusola, J. Ballantyne, Messenger RNA profiling: a prototype method to supplant conventional methods for body fluid identification, Forensic Sci.Int. 135 (2003) 85-96.
[125] J. Juusola, J. Ballantyne, Multiplex mRNA profiling for the identification of body fluids, Forensic Sci.Int. 152 (2005) 1-12.
[126] J. Juusola, J. Ballantyne, mRNA profiling for body fluid identification by multiplex quantitative RT-PCR, J.Forensic Sci. 52 (2007) 1252-1262.
[127] Y. Matsubara, H. Ikeda, H. Endo, K. Narisawa, Dried blood spot on filter paper as a source of mRNA, Nucleic Acids Res. 20 (1992) 1998.
[128] T.W. Phang, C.Y. Shi, J.N. Chia, C.N. Ong, Amplification of cDNA via RT-PCR using RNA extracted from postmortem tissues, J.Forensic Sci. 39 (1994) 1275-1279.
[129] M. Setzer, J. Juusola, J. Ballantyne, Recovery and stability of RNA in vaginal swabs and blood, semen, and saliva stains, J.Forensic Sci. 53 (2008) 296-305.
[130] D. Zubakov, M. Kokshoorn, A. Kloosterman, M. Kayser, New markers for old stains: stable mRNA markers for blood and saliva identification from up to 16-year-old stains, Int.J.Legal Med. 123 (2009) 71-74.
[131] P. Chomczynski, N. Sacchi, Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction, Anal.Biochem. 162 (1987) 156-159.
[132] E. Hanson, C. Haas, R. Jucker, J. Ballantyne, Specific and sensitive mRNA biomarkers for the identification of skin in 'touch DNA' evidence, Forensic Sci.Int.Genet. 6 (2012) 548-558.
[133] E.K. Hanson, J. Ballantyne, Highly specific mRNA biomarkers for the identification of vaginal secretions in sexual assault investigations, Sci.Justice. 53 (2013) 14-22.
[134] E. Hanson, K. Rekab, J. Ballantyne, Binary logistic regression models enable miRNA profiling to provide accurate identification of forensically relevant body fluids and tissues, Forensic Sci.Int.Genet.Supp.Ser. 4 (2013) e127-e128.
[135] D. Zubakov, A.W. Boersma, Y. Choi, P.F. van Kuijk, E.A. Wiemer, M. Kayser, MicroRNA markers for forensic body fluid identification obtained from microarray screening and quantitative RT-PCR confirmation, Int.J.Legal Med. 124 (2010) 217-226.

VI. DISSEMINATION OF RESEARCH FINDINGS

A. Publications

Due to the challenges in identifying suitable RNA candidates for the determination of the biological age of a bloodstain donor, no publications have occurred for the current study at this time. We are currently in the process of further validating the elderly mRNA assay (including logR model) and a manuscript summarizing the development and validation of this assay will be prepared upon completion of that work. Furthermore, we are hopeful that the additional recently identified potential mRNA and miRNA biomarkers will be successful in predicting the age of the donor. If successful, publications summarizing the development and validation of these assays will be produced.

B. Presentations

1. A presentation entitled "Identification of Fetal Blood Using Developmentally Regulated Gamma Hemoglobin Isoforms" was given by Jack Ballantyne as part of the Forensic Technology Center of Excellence's Rapid Biological Screening and Analysis Methodologies for Improving Throughput ("Field") Technology Transition Workshop (TTW). The presentation was taped previously and then aired via the web. A live question and answer session followed the webinar. August 2011.
2. Albornoz, A., Hanson, E. and Ballantyne J. Predicting the Biological Age of a Bloodstain Donor by mRNA Profiling. 9th Annual Showcase of Undergraduate Research Excellence. University of Central Florida. April 2012.

VII. ACKNOWLEDGEMENTS

The authors would like to acknowledge Dr. Kamel Rekab (University of Missouri Kansas City (UMKC) for his collaboration and logistic regression analysis and model development. The authors would like to acknowledge all laboratory personnel (past and present) who have contributed to this project: Erin Hanson, Tarah Marks, Mohid Mirza, Laura Guzman, Sami Nelson, Mindy Setzer, Ariana Albornoz, Miriannette Gayoso, Tisha Choudhury and Michelle Alvarez.

VIII. RESPONSE TO REVIEWERS

Reviewer 1:

Review of project technical report/deliverables

Award number: 2009-DN-BX-K179
Project title: The Biological Age of a Bloodstain Donor

Practitioner need addressed by project: Identification of RNA markers for predicting the biological age of a forensic bloodstain sample donor

Reviewer provides information below:

1. **Substantive Quality**
   - What are the significant findings of the research?

The authors previously proposed forensic mRNA markers – first of a kind - for distinguishing neonatal samples from other age groups using novel hemoglobin G (HBG) mRNA isoforms. In the current project they performed a whole genome scan for novel markers discriminating between the age categories that cover the entire lifespan: newborn/infant (birth - < 2 years); toddler (2-6 years); child (7-12 years); adolescent (13 – 17 years); adult (18 – 45 years); mature adult (46 – 65 years); and elderly (≥ 65 years). In addition to mRNA profiling performed by state of the art RNA-Seq technology the authors also explored the whole variety of human miRNAome aiming at identification of respective markers. To my knowledge, the latter part of the project is unique; no previous attempts of identification of forensic miRNA age markers were ever undertaken. The second approach employed by the authors to identify age markers involved targeting specific candidate genes selected by a priori understanding of the physiology and biochemistry of human development. Over 500 mRNA and 1,100 miRNA age marker candidates were evaluated. Eventually, a set of 11 novel mRNA gene candidates, that along with two housekeeping genes (B2M, S15) and our previously identified newborn markers (HBG1n and HBG2n), permit an identification of an individual as belonging to one of 5 major age group classifications. For the analysis of these ~15 mRNA markers four qPCR assays were developed. Logistic regression analysis was used for predictions of the biological ages in three of the four assays. Along with mRNA candidates, 36 miRNAs were identified as potential age markers, their evaluation is ongoing.

In general, the authors presented one of the first set of RNA age markers that was developed and validated for forensic purposes. The results obtained by the authors are essentially new and highly relevant for forensic age estimation of bloodstain donors.

Response: None needed

- Are the findings supported by the research? Was the methodology appropriate and sound?

All the conclusions in this report are supported by solid experimental evidences. The design of the project was elaborate and intelligent. Methodology used by the authors was adequate and appropriate for the tasks of the study.

Response: None needed

- Does the Executive Summary adequately describe the full report?

The Executive Summary describes the report correctly and comprehensively.

Response: None needed

- Is the report well-written in terms of style, organization, and format?

The report is very well written in respect of all aspects.

Response: None needed

- Classify the overall quality as one of the following: Poor, Fair, Good, Excellent.

I believe that the overall quality of the project is Good.

Response: None needed

2. Implications of the Research

- Do the findings make a significant contribution to existing knowledge in the area?

The findings of the authors described in this report significantly extend the existing knowledge in the area in two major respects:

1. *the authors showed that relatively accurate age prediction can be achieved by the analysis of rather limited number of mRNA markers*
2. *they obtained one of the first evidence upon the usefulness of miRNA markers for forensic age estimation*

Response: None needed

- What are the implications, if any, for further research, program development, and evaluation efforts?

The work performed by the authors should promote further forensic studies of age markers, including mRNA and miRNA, as well as other types of molecular markers. Set of mRNA markers and qPCR assays developed by the authors should be evaluated internationally and, when successful, it may be implemented in forensic casework.

Response: None needed

3. Relevance

- Summarize your overall rating of the report's/deliverable's relevance for policy or practice as one of the following: Poor, Fair, Good, Excellent

In my opinion practical relevance of the reported project is Excellent. Age estimation with molecular markers is becoming a hot topic in forensic science.

Response: None needed

- To whom would this report/deliverable be of greatest interest?

Forensic scientists and practioners is the target group that should be highly interested in the results of this research.

Response: None needed

- Would you recommend this information for publication?

Indeed, the results obtained by the authors are valuable from both theoretical and practical points of view and should be shared with forensic community.

Response: None needed

- What specific dissemination vehicles would be particularly appropriate for publicizing the research (e.g., conferences, scientific literature, etc.)?

Specialized forensic journals such as Forensic Science International: Genetics should be primary media for dissemination of results. The results should also be disseminated through scientific conferences.

Response: None needed

- How well does this report/deliverable address the practitioner need as stated above?

Before implementing the results of current study in practical forensic casework, they need to be validated by independent forensic labs. The authors themselves previously organized several EDNAP trials of mRNA markers for body fluid identification that greatly promoted their practical application. I suppose that the age mRNA markers can be evaluated via EDNAP exercises as well.

Response: None needed

4. Recommended revisions

Describe fully any revisions or changes that should be made to improve the quality of the report or increase its usefulness. Your anonymous comments and suggestions will be forwarded to the author of the report, who will make appropriate revisions in order to improve the report.

I have only few comments and suggestions:

1. *There are several published papers on age-related mRNA markers in human blood, mostly from general biological field of research. It would interesting to learn what is the overlap between the candidate age markers identified in current project and by other authors. My impression is that the overlap is very poor, which is not surprising in this sort of studies, but has to be mentioned and, if possible, explained in the report.*

Response: The RNA age markers we identified have not, to our knowledge, been reported elsewhere as being biomarkers of age in dried bloodstains. Other age biomarkers have been reported mainly in non-self-renewing tissue types but haven't included our identified biomarkers. If they had been previously identified we would indeed have mentioned this in the report's review of literature.

2. *The sample size seems to be very small, especially in discovery set used for RNA-Seq experiments. Also, validation samples size is not big enough. From my experience, many promising candidates looking very well in dozens of samples may easily became barely significant when evaluated in hundreds or thousands of samples*

   Response: While the number of samples used in the RNA-Seq data may appear quite small, the purpose of that experiment was initial discovery and therefore two donors at each age group was sufficient. The cost associated with the analysis of this number of samples was extremely high ($34,000). Therefore we were working within the limits of our budget at the time of analysis. Additionally, this data still provided an initial assessment of gene expression in various age groups which, after appropriate screening, was sufficient to identify potential promising candidates. Any candidate identification from this data was followed up with additional testing with qPCR assays. Therefore the RNA-Seq data provided useful information for the project to proceed. Personal or affordable NGS instruments were not available when this analysis was performed as they are today.

   The authors clearly state in the report that "the assays developed in the current work will all require additional validation before they could be used routinely for age determination". Like any NIJ study, additional samples would indeed need to be tested to increase sample size and further validate the developed assays.

3. *The authors used age categories (newborn, toddler, child, adolescent, adult, mature adult, and elderly) of unequal size, justifying this by the following: "generally correlate with known psychosocial and biological changes known to occur throughout the human development life cycle". However, there are only few quantum changes in human development that really occur in some particular narrow ages (e.g. puberty or menopause), whereas the most of aging processes are gradual and continuous. Therefore, I wonder if using regression (linear or non-linear), instead of categorical, approach for the discovery of age markers would bring more potential candidates?*

   Response: As was stated in the report: "These initial age sub-categories represent the developmental phases of the human life cycle and were used heuristically in the event that distinct biomarkers for each age group could be identified. However, we realize that these designations may not necessarily represent absolutely distinct 'gene expression boundaries' and the precise ages that differentiate the age groups may be different to the ones we chose on an ad hoc basis. Therefore, for interpretation of any obtained expression data throughout this project we were careful not to rely strictly on these classifications and to evaluate the data in a broader non-biased sense without

pre-judging to which category any sample (that was close to an age classification boundary) belonged to."

However, the initial assays developed were able to provide resolution with a reasonable approximation to the categories originally proposed. The use of logistic regression proved to be useful in the development of our age models. As with any assay development alternative statistical approaches can be used. The use of logistic regression proved to be useful in the development of our age models. However it certainly is possible that linear or non-linear regression analysis could have yielded more potential candidates.

Reviewer 2:

1. Substantive Quality

a. What are the significant findings of the research?
The research has begun to identify genetic age-correlated markers that could be useful in narrowing the age group of the donor of a bloodstain.

Response: None needed

b. Are the findings supported by the research? Was the methodology appropriate and sound?
The research is comprehensive and complex, but thoroughly carried out with appropriate *a priori* assumptions. The findings are supported by the research.

Response: None needed

c. Does the Executive Summary adequately describe the full report?
The Executive Summary is an excellent summary of the full report.

Response: None needed

d. Is the report well-written in terms of style, organization, and format?
The report is extremely well written, concise, and replete with adequate documentation, figures, tables, and graphs. The format is excellent, easy to follow, and the methods are made understandable.

Response: None needed

e. Classify the overall quality as one of the following: Poor, Fair, Good, Excellent.
The overall quality of the report is Excellent.

Response: None needed

2. Implications of the Research

- Do the findings make a significant contribution to existing knowledge in the area?
  The findings are preliminary (several assays developed) as to which markers are best for categorizing age, but this is basic research. Although the research so far is most successful in identifying markers that define the extreme age ranges of humans (infants and the elderly), perhaps the most important contribution is that the research makes it clear that it will be difficult to refine this system further, much less make it easy for the average forensic lab to apply the tools to define the age of blood donor.

  Response: The authors agree that this work implies that age biomarkers can perhaps be identified but the resolution obtained will be difficult to go beyond the broad categories of human development (newborn, elederly etc). However with the massively parallel analytical capabilities of the next generation sequencing technologies (NGS) about to further revolutionize forensic biology, it should be a facile task to incorporate a number of these age biomarkers into the 'external visible trait assays' that are being incorporated into these comprehensive forensic genomics assays.

- What are the implications, if any, for further research, program development, and evaluation efforts?
  The implications of the project results are that 1) further funding of this work should be considered carefully in light of its limited utility (especially if more refined age categories cannot be elucidated), and 2) that the true application of these complex assays may not be useful for an average forensic lab, but may be limited to academic laboratories.

  Response: We disagree that there is limited utility for the average forensic laboratory and that only two age groups are identified by the developed assays. The workflows provided in Figures 2 and 3 demonstrate a potential for determination of 5 different age groups. Additionally, physical characteristics, including age, are being investigated for routine use in forensic analysis – hair color, eye color, skin pigmentation, height, weight, etc. This would permit the development of a "genetic eyewitness" and not rely on eyewitness testimony or provide valuable probative information when no suspect is identified and eyewitness accounts are not available. While the routine use of this analysis awaits development of possible massively parallel assays so that all biomarkers could be investigated at one time (reducing the time and cost of analysis), this is eminently feasible with the increasing use of NGS technology (also see previous response above)

3. Relevance

- Summarize your overall rating of the report's/deliverable's relevance for policy or practice as one of the following: Poor, Fair, Good, Excellent

The report's relevance is excellent as it clarifies what the limitations of this research are. There are no true deliverables in that the forensic community cannot benefit in a concrete way from these methods.

Response: It is unclear why the reviewer thinks that the forensic community could not benefit in a concrete way from these methods. They provide a means by which to determine the biological age from a bloodstain - contributing to the determination of physical characteristics of the donor of a biological sample. So unless one perceives physical characteristics to be a useless tool for forensic scientists, then it is unclear why a determination of biological age would not benefit the forensic community.

The true number of cases in which these methods could actually be applied is very small, and in fact the most relevant populations to which they *should* be applied are young to middle adults, a population for which these assays do not appear to work well at this time. For example, for identifying the age of a criminal perpetrator, most of whom fall into the mid-adult age range, the assays are not helpful. In terms of locating an unknown blood stain of interest, it is difficult to imagine a scenario in which a kidnap victim or elderly missing person's age would not be known. In those scenarios, a normal STR profile would be obtained to compare with family members.

Response: We need as a community to start moving away from the typical refrain that something is not worth doing because the number of cases may be small. What may be a small number in a local sense may be a substantial number in a National sense. Age is a type of information that probably comes under the umbrella of forensic intelligence used by investigators to narrow potential leads and exclude false leads.

- To whom would this report/deliverable be of greatest interest?
  Academic/pharmaceutical/medical research labs that study longitudinal gene expression patterns over the life-span of an individual.

  Response: See above comments regarding relevance to the operational forensic community.

- Would you recommend this information for publication?
  The information should be published as basic research, not necessarily in forensic venues.

  Response: The authors disagree that this would not be suited for publication in forensic venues as it has relevance to forensic casework. If something contributes to a body of knowledge, as this arguably does in the area of forensic science, then the relevant professional forensic journals would seem to be the most appropriate place to publish the results.

- What specific dissemination vehicles would be particularly appropriate for publicizing the research (e.g., conferences, scientific literature, etc.)?
  The scientific literature (when the project has reached a suitable state of completion) is the best venue for dissemination.

  Response: None needed

- How well does this report/deliverable address the practitioner need as stated above?
  This project does not address a practitioner need in a significant way. The likelihood of any validated assay being used routinely is small. Keeping in mind that all laboratories must routinely proficiency test (which is expensive) in applied methods, the low level of use will prohibit adoption in the average lab. Finally, the need for identification of the age of a bloodstain donor is small.

  Response: The authors again disagree with this opinion. As stated above, with the eminent development of NGS based assays with the capacity to evaluate hundreds to thousands of biomarkers in a single assay, there could indeed be a place for age determination in such assays. While this information may not be needed in every case, it would be probative information that could be useful in investigations.

4. Recommended revisions

Describe fully any revisions or changes that should be made to improve the quality of the report or increase its usefulness. Your anonymous comments and suggestions will be forwarded to the author of the report, who will make appropriate revisions in order to improve the report.

There are no recommended revisions of changes.

Response: None needed