International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

http://gssrr.org/index.php?journal=JournalOfBasicAndApplied

Use of mRNA Markers for Age Prediction in Healthy and Unhealthy Individuals of Indian Subcontinent

Jawad Salehi[a]*, Lamis Abdelaal[b], Rania Gomaa[c]

[a,b,c]*College of Biotechnology, University of Modern Sciences, United Arab Emirates*

[b]*Forensic Medicine and Clinical Toxicology Department, Faculty of Medicine, Cairo University, Egypt*

[c]*Forensic Medicine and Clinical Toxicology Department, Faculty of Medicine, Alexandria University, Egypt*

[a]*Email: j.salehi@ums.ae,* [b]*Email: l.abdelaal@ums.ae,* [c]*Email: r.gomaa@ums.ae*

Abstract

Current age estimation techniques rely on morphological features such as teeth eruption, wear and tear pattern, and fusion of skeletal bones at different stages of life. These techniques require the presence of complete or partial human body and sometimes are difficult to perform on living individuals. Age estimation using molecular markers can provide better accuracy and can be performed with very small amount of biological sample found at the crime scene. Recently, mRNA, T-cell specific DNA rearrangements and mitochondrial DNA have been tested for estimating the age, and mRNA was found to be most age correlated. This research aims to use mRNA as a genetic marker to predict the age of Indian subcontinent population, both healthy and unhealthy individuals, by correlating the expression patterns within different age groups. This research provides an insight about the potentials of using mRNA markers to predict biological or chronological age by testing them on both healthy and unhealthy individuals. This study compared the predictive powers of three genes (NRCAM, CFH, and SLC16A10) against healthy individuals and unhealthy individuals. The results showed no significant difference in age prediction between healthy and unhealthy individuals. In order to further confirm the results future studies should include samples from various populations with large sample size to produce more statistically sound data which can further help improve age prediction process in forensic investigations.

Keywords: Forensic Science; Age estimation; mRNA.

---

* Corresponding author.

1. Introduction

Estimation of phenotypic characteristics has high importance in forensic investigations to create biological profile of victim and/or perpetrator. Among phenotypic characteristics gender and age estimation has high significance. Currently, there are many techniques that can measure age with varying degrees of accuracy. These techniques range from dental and skeletal morphological analysis to molecular level studies. Morphological techniques predict age with wide range (e.g. +/- 10 years) and require the presence of teeth or skeletal remains. These techniques also require the presence of high expertise and are sometimes subjective. Molecular techniques do not require presence of full body and can estimate age from as little as a blood stain. Recently, many biological markers have been investigated for their use as age-predictive markers. These markers include telomere repeats, DNA methylation, T-cell DNA, mitochondria and mRNA. All of these markers are able to predict age with varying accuracies [1].

Genes tend to up regulate or down regulate as the individual ages. These changes of gene expression can be used as a marker to predict age. As changes in gene expressions can be correlated with age, we can screen for genes in large sample size with all the age groups and select for genes that show correlation with age and have variation throughout the age ranges.

Alvarez and his colleagues conducted the first study that used mRNA for newborn identification. mRNA marker from gamma hemoglobin gene transcript was investigated and showed age related expression and was only expressed in newborn and thus will be useful to confirm if the source of blood stain is newborn or not [2].

Recently, Zubakov and his colleagues (2016) screened 222 age related genes using microarray technology. From these they selected top 9 most age correlated mRNA markers. These markers were then validated on 267 individuals (142 females and 125 males) with ages ranging from 22-84 years. Results have shown ability to estimate age with +/- 11.595 years using the constructed age prediction model [1]. Two of the mRNA markers were also able to predict age groups (22-40, 41-60, 61-84) using the area under curve (AUC) approach. The authors concluded that DNA methylation has highest age predicting power followed by mRNA and thus further researches are required using them as markers [1].

Even though numerous studies have been performed on age-related markers none of these studies have tested the prediction powers of these markers on unhealthy individuals. That is if these markers are able to predict chronological age or do they estimate biological age of an individual. The present study is first to employ mRNA age-predictive marker on both healthy and unhealthy individuals in order to assess their predictive power on a population group and their ability to predict chronological or biological age [1].

2. Materials and Methods

2.1. Primers

Primer sequences were retrieved from “Primer Bank” https://pga.mgh.harvard.edu/primerbank/. Primers were then checked for self-complementarity and potential hair pin formation using “Oligonucleotide Properties

Calculator" http://biotools.nubic.northwestern.edu/OligoCalc.html. Primer sequences were also validated for specificity using NCBI Primer blast [Table 1].

Table 1: Primer details for NRCAM, SLC16A10, CFH and GAPDH.

Primer	Annealing temperature (^{o}C)	Forward (5'-3') sequence	Reverse (5'-3') sequence
NRCAM	62	TCCAACCATCACCCAACAGTC	TGAGTCCCATTACGGGTCCAG
SLC16A10	61	TAGTCAGCGTCTTCACAGACC	GAGCCCAACAAATCCAACAGC
CFH	58	GTGAAGTGTTTACCAGTGACAGC	AACCGTACTGCTTGTCCAAAA
GAPDH	59	ACAACTTTGGTATCGTGGAAGG	GCCATCACGCCACAGTTTC

4.2. Sample collection

Peripheral blood samples were collected from 29 individuals living in UAE (19 healthy and 10 unhealthy) from a Medical center. Informed consent was taken from the volunteers prior to sampling. All samples were from male individuals of age range 20-54 years old. Individuals suffering from cardio vascular diseases and diabetes were considered as unhealthy in the current study. The samples represent Pakistani, Nepali, Bangladeshi and Indian nationality (Indian Subcontinent).

4.3. RNA extraction

RNA was extracted using NorgenBioteks "Total RNA Purification" kit and manufacturer's instructions were followed. The samples were then stored at -80 ^{o}C.

4.4. DNase 1 Treatment

DNase 1 used in this research was obtained from New England Biolabs. Extracted RNA was quantified using SpectraX microplate reader and then 10 ug of RNA was resuspended in 1 X DNase I reaction buffer (New England Biolabs) for a total volume of 100 uL. Then 2 units of DNase I was added to the mixture. The mixture was then incubated at 37^{o}C for 10 minutes. After the incubation 0.5 M of EDTA was added to the mixture in order to protect RNA degradation from high heat. The mixture was then incubated at 75^{o}C for 10 minutes to inactivate DNase enzyme.

4.5. First Strand cDNA Synthesis

First strand cDNA synthesis was performed using kit from New England Biolabs. RNA was converted to cDNA using OligoDT primers and manufacturer's instruction was followed. The converted cDNA samples were stored at -20 ^{o}C.

4.6. PCR optimization

Primers were optimized using 10X PCR master mix from NorgenBiotek. Gradient PCR was used in order to determine annealing temperature that produced better results [Table 2].

Table 2: Polymerase chain reaction conditions.

Stage	Temperature (^oC)	Time	Cycles
Initial Denaturation	95	5 minutes	1
Denaturation	95	30 seconds	
Annealing	58-62	30 seconds	35
Extension	72	30 seconds	
Final Extension	72	7 minutes	1

The PCR reaction was prepared for total volume of 20 uL. With 10 uL of master mix, 1 uL of forward primer, 1 uL of reverse primer, 7 uL of nuclease-free water and 1 uL of sample DNA in each reaction tube.

4.7. Real time PCR

Real time PCR was performed to quantify the gene expression of selected genes. This was done using SYBR-green master mix from BioRad [Table 3].

Table 3: Real time PCR conditions

Stage	Temperature (^oC)	Time	Cycles
Initial Denaturation	95	5 minutes	1
Denaturation	95	30 seconds	
Annealing	58-62	30 seconds	35
Melt Curve	65-95 (0.5^oC increment, 0.05 seconds hold)	7 minutes	1

The Real time PCR reaction was prepared for total volume of 20 uL. With 10 uL of SYBR green master mix, 1 uL of forward primer, 1 uL of reverse primer, 7 uL of nuclease-free water and 1 uL of sample DNA in each reaction tube.

4.8. Statistical analysis

Statistical analysis was performed using Microsoft Excel 2013. A multiple regression was performed to create an age prediction model from the expression of NRCAM, SLC16A10 and CFH genes. And the following equation was used as the basis of this multiple regression analysis:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + e_i$$

Where β_0 is the constant and β_1 till β_p is the coefficients for the explanatory variables (x values).

5. Results

5.1. RNA extraction

RNA extraction was confirmed by running the sample on 1.5 % agarose gel. In the presence of total RNA 2 bands were present on agarose gel corresponding to 28S and 18S rRNA. However, another band (first one) was also present which was of genomic DNA. RNA quality was determined using spectrophotometer at wavelength 260 and 280. A ratio of 260/280 around 2 is considered pure RNA and lower than that can be due to the presence of proteins and phenols. In this study all the RNA samples showed ratio close to 2.

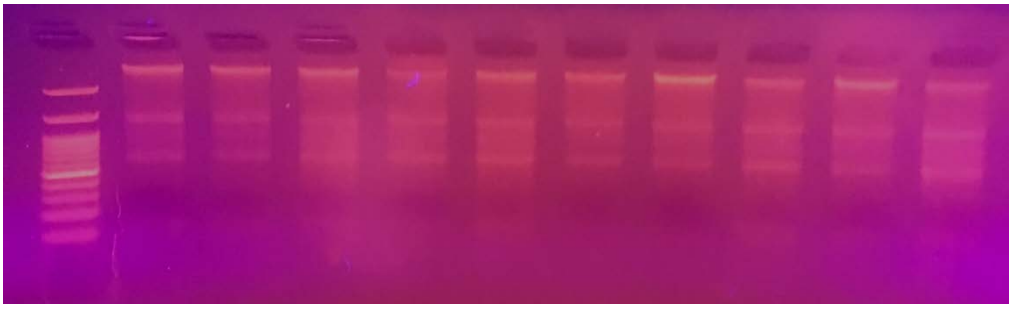

Figure 1: Gel Electrophoresis of Extracted RNA showing: gDNA, 28S and 18S rRNA

5.2. First strand cDNA synthesis

In order to check if there was any gDNA contamination present a PCR reaction was carried out using primer specific to NRCAM gene (specific to variant 6 of the transcript). In the presence of gDNA 2 bands were produced of size 1100 bp and 123 bp and in the presence of cDNA only a single band of 123 bp was produced.

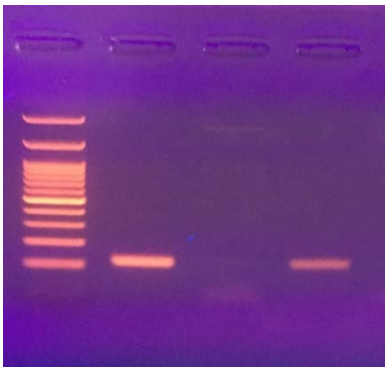

Figure 2: Presence of 1100 bp product in sample with gDNA contamination (lane 3). Lane 1 and 2 show pure cDNA with 123 bp amplification

5.3. Real time PCR

After primer optimization cDNA was quantified using real time PCR and SYBR-green mastermix. During each set of runs a negative control was also included which showed no amplification. Primer specificity was also checked by studying the melt curves during each run. The melt curves obtained in this research had single peak confirming the presence of one product.

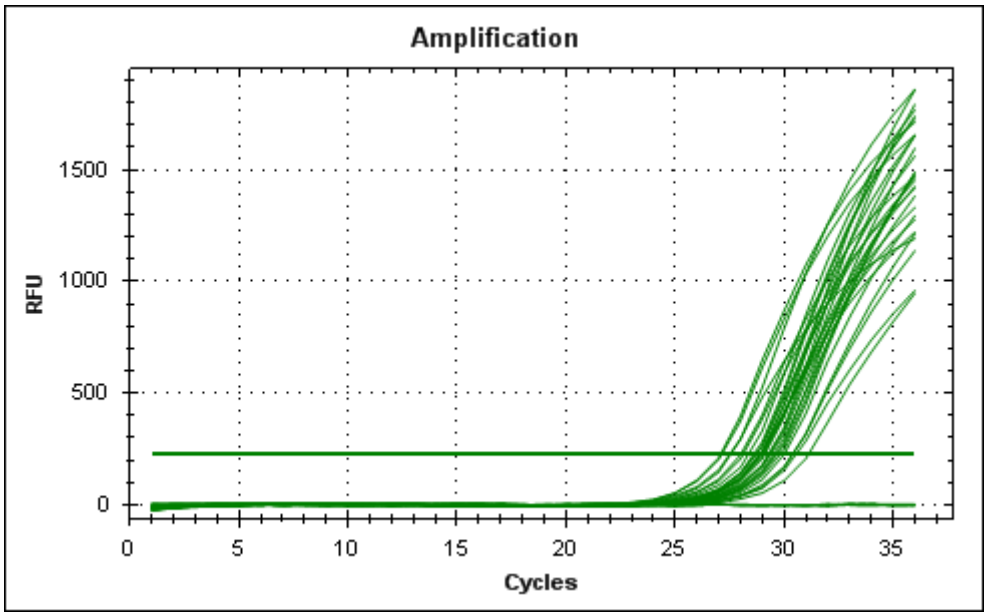

Figure 3: Quantification curve showing no amplification of negative control

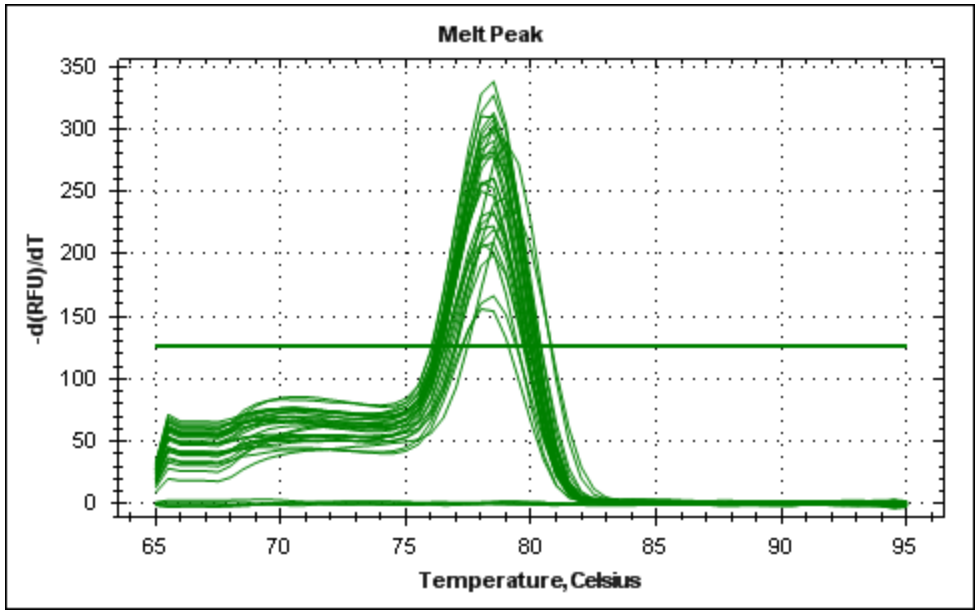

Figure 4: Melt curve with single peak showing absence of non-specific binding of primers

All samples showed amplification with all four primers except for four samples which were of highest age

among the samples. These four samples failed to amplify with NRCAM primer. Over all NRCAM primers produced highest Ct value when compared to other three primers. The Ct value produced with NRCAM primer was in 31-35 range. Whereas, Ct values for other primers were similar with range of 24-30.

NRCAM showed a negative correlation with age and its expression was down regulated with age. Whereas, both SLC and CFH showed positive correlation with age and their expressions were upregulated with age.

5.4. Statistical analysis

The prediction model was built using data from 14 samples of healthy individuals. And this model was tested on 5 healthy samples and further investigated on 10 unhealthy samples to study the effect of health on the prediction powers of markers.

Multiple linear regression was performed on the real time PCR data using Microsoft Excel with Analysis ToolPak add-in. The Ct values of all the three genes (NRCAM, CFH, SLC16A10) were transformed to log with base 2 and then normalized using GAPDH as the reference gene following simple Δ Ct method, where Δ Ct = Ct target – Ct reference. The resulting values were then used as independent variables (as X values) and the known age of the individuals were used as dependent variable (as y values). Multiple regression model was generated with r^2=0.5 and standard deviation of +/-6.7. ANOVA test showed that there is only 3.9% chance that the outcome of regression is by chance and not by the correlation between the variables. The p-values for all the three genes was also calculated with NRCAM showing the lowest p-value of 0.08. SLC and CFH had higher p-values of 0.21 and 0.78 respectively.

Residual output was also produced which showed the difference between the actual and predicted dependent variable. A scatter plot was generated using the residual data to check if it produced any pattern. The residuals were normally distributed across zero with no pattern.

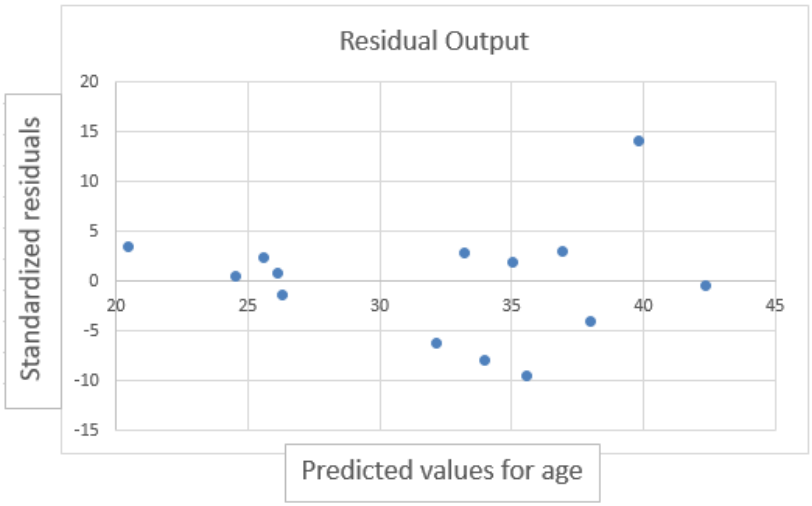

Figure 5: Residual Plot between Standardized and predicted values

The predictions made using the model was mostly accurate with low deviation from an actual age. However, predictions for few samples in both healthy and unhealthy showed large deviation from their actual age; for example, sample 5 and 10 for unhealthy set [Table 4] and sample 2 for healthy set [Table 5].

Table 4: Actual and predicted age of unhealthy individuals.

Unhealthy		
No.	Actual Age	Predicted Age
1	33	37
2	38	40
3	32	43
4	34	38
5	20	43
6	30	32
7	47	50
8	38	38
9	37	43
10	49	59

Table 5: Actual and predicted age of healthy individuals.

Healthy		
No.	Actual Age	Predicted Age
1	24	20
2	26	34
3	27	26
4	28	25
5	25	26

5. Discussion

All samples showed amplification with all four primers except for four samples which were of highest age among the samples. These four samples failed to amplify with NRCAM primer. Over all NRCAM primers produced highest Ct value when compared to other three primers.

NRCAM is a neuronal gene which is responsible for the production of NRCAM protein which is a cell adhesion molecule. NRCAM is mostly expressed in brain and spinal cord where it is involved in the transmission of neural signals. NRCAM also acts as a receptor for numerous neuronal molecules [3]. NRCAM protein is present in blood in negligible amount and therefore the cDNA produced are in low copy number and thus produced high Ct values consistently in all samples. NRCAM gene is shown to be negatively correlated with age [1] and therefore the old aged samples did not show any amplification with NRCAM specific primers. In this study

NRCAM showed high correlation with age when compared to the other genes (CFH and SLC16A10) and the results were consistent with the previous study [1].

CFH is the member of Complement activation gene cluster family which produces the CFH protein and secretes it in the blood stream. CFH protein is involved in the defense against microbial infections but is extensively studied for its involvement in Age-related macular degeneration (AMD). It has been shown in some studies that presence of certain SNPs in CFH gene can make an individual more susceptible to AMD development in later stages of life [4]. One study [1] showed that CFH is positively correlated with age and that its expression is increased as the individual ages which explains high rates of AMD in older individuals compared to their younger counterparts. In this study CFH was shown to be positively correlated with age and the results were consistent with other study [1].

SLC16A10 gene is a member of solute carrier (SLC) proteins. These proteins are responsible in transportation of various components through and across cells. SLC16A10 is especially involved in the transport of thiamine pyrophosphate in the mitochondria. This compound is important for energy production and has role in Krebs cycle [5]. And in previous study [1] was able to show that this gene is negatively correlated with age, which explains fall in mitochondrial efficiency in older aged individuals. However, in this study SLC16A10 was shown to be upregulated with age instead of being down regulated. This contradiction in the results is due to the usage of one region of the gene was as primer binding site to quantify its expression. On the other hand [1] used multiple Taqman probes specific to multiple sites spread along the gene. This gave them higher resolution and accuracy in quantifying its expression than the approach used in this study.

Among the three genes tested in this study NRCAM showed higher correlation with age and this result was consistent with previous study which also showed NRCAM having highest correlation with age [1].

The prediction model created using multiple linear regression predicted age with little deviation from actual except for some samples. This model shows how promising mRNA is in predicting chronological age of an individual given more studies are conducted to search for more markers that are highly correlated with age and are less affected by individuals health.

6. Conclusion and Limitations

This study compared the predictive powers of three genes (NRCAM, CFH, and SLC16A10) against healthy individuals and unhealthy individuals. Previously, the three genes showed promising results in predicting age of healthy individuals with low deviation from actual age. Therefore, this study was conducted to test if the proposed mRNA markers predict chronological age or biological age of an individual. And thus unhealthy samples of known age were also tested using the prediction model built using these genes. The results showed no significant difference in age prediction between healthy and unhealthy individuals.

Major limitation while conducting this research was the degradation of RNA while transporting the blood sample from medical laboratories to the research facilities. However, the samples were transported in RNase free EDTA tubes placed in ice cold box which slowed the degradation process.

7. Recommendations

In order to further confirm the results future studies should include samples from various populations with large sample size to produce more statistically sound data which can further help improve age prediction process in forensic investigations. Also genome wide association studies (GWAS) can further help in identifying more candidate mRNA markers for age prediction.

Acknowledgement

The authors are thankfull to the University of Modern Sciences for providing facilities to conduct this reseach and for funding this research.

References

[1] D. Zubakov, F. Liu, I. Kokmeijer, Y. Choi, J. van Meurs, W. van IJcken, A. Uitterlinden, A. Hofman, L. Broer, C. van Duijn, J. Lewin and M. Kayser. “Human age estimation from blood using mRNA, DNA methylation, DNA rearrangement, and telomere length”. Forensic Science International: Genetics, 24, pp.33-43, 2016.

[2] M. Alvarez and J. Ballantyne. “The identification of newborns using messenger RNA profiling analysis.” Analytical Biochemistry, 357(1), pp.21-34, 2006.

[3] T. Marui, I Funatogawa, S. Koishi, K. Yamamoto, H. Matsumoto, O. Hashimoto, E. Nanba, H. Nishida, T. Sugiyama, K. Kasai, K. Watanabe, Y. Kano and N. Kato. “Association of the neuronal cell adhesion molecule (NRCAM) gene variants with autism.” The International Journal of Neuropsychopharmacology, 12, pp.1-10, 2009.

[4] P. Francis, D. Schultz, S. Hamon, J. Ott, R. Weleber and M. Klein. “Haplotypes in the Complement Factor H (CFH) Gene: Associations with Drusen and Advanced Age-Related Macular Degeneration.” PLoS ONE, 2(11), p.e1197, 2007

[5] V. Iacobazzi, M. Ventura, G. Fiermonte, G. Prezioso, M. Rocchi and F. Palmieri. “Genomic organization and mapping of the gene (SLC25A19) encoding the human mitochondrial deoxynucleotide carrier (DNC).” Cytogenetic and Genome Research, 93(1-2), pp.40-42, 2001.