Annals of Medical Research

DOI: 10.5455/annalsmedres.2019.09.578 2019;26(12):3032-6

Original Article

A bioinformatical approach to the pathogenesis of Fragile X premutation carriers

Elif Pala

SANKO University, Faculty of Medicine, Department of Medical Biology, Gaziantep, Turkey

Abstract

Aim: The common cause of hereditary mental retardation is Fragile X Syndrome (FXS). This disease occurs when the trinucleotide repeat number (CGG)n in the promoter region of the Fragile X mental retardation 1 (FMR1) gene located in Xq27.3 is increased. Fragile X-associated tremor/ataxia syndrome (FXTAS) is a neurodegenerative disease caused by CGG repeat increase to 55-200 in the FMR1 gene. Although regulation disorders or mutations in FMR1 seem to be responsible for the pathogenesis of FXS and FXTAS, the broad phenotypic spectrum suggests that there should be some other potential genes involved. In this study, it was aimed to determine the differentially expressed genes of FMR1 premutation carriers and healthy controls by using bioinformatics techniques.

Material and Methods: Gene expression profiles were downloaded from the Gene Expression Omnibus (GEO) database, which were analyzed using the Principal Component Analysis (PCA)-based unsupervised feature extraction (FE).

Results: The set of 14 genes were identified that could successfully discriminate fragile X premutation carriers and healthy controls, and the majority of these genes were long non-coding RNAs (lncRNA).

Conclusion: Although the results of our study should be supported by extended experimental researches, these genes have the potential to be used as biomarkers and therapeutic targets.

Keywords: Fragile X; principal component analysis; unsupervised feature extraction; biomarker; lncRNA

INTRODUCTION

Trinucleotide repeat disorders are a set of genetic diseases that are often associated with neurological diseases caused by the increase in the repetition of the trinucleotide in particular genes (1). Fragile X Syndrome (FXS) is the most common cause of hereditary mental retardation and is the second common cause of mental retardation after Down syndrome (2,3). The Fragile X Mental Retardation 1 (FMR1) gene in the q27.3 region of the X chromosome contains high polymorphic CGG repeats in the 5′ untranslated region (5′-UTR) (2-4). In the general population, the repeat of the FMR1 CGG ranges from 5 to 55 copies, and the most common allele is 30 repeats (5). FMR1 CGG repeats ranges from 55 to 200 are classified as premutation alleles, while more than 200 repeats characterize full mutation alleles. In entirely mutated alleles, the gene is usually hypermethylated and silenced, resulting in FXS (6). Premutation carriers have normal or decreased FMR1 protein, and mRNA levels are 2-8 times higher than normal alleles. These individuals are often asymptomatic but are at risk of having affected children because CGG repeat numbers are unstable and tend to increase with each cell division (7). In studies, premature alleles have been associated with various disorders such as Fragile X Tremor/Ataxia Syndrome (FXTAS) (8), premature ovarian failure (9), thyroid dysfunction, hypertension, fibromyalgia and chronic muscle pain (10,11). Although regulation disorders or mutations in FMR1 seem to be responsible for the pathogenesis of FXS and FXTAS, individuals without mutations in the FMR1 gene that carry phenotypic symptoms of FXS are available (12,13).

Also, only about one-third of male premutation carriers develop FXTAS (14). All these knowledge suggest that, in addition to FMR1, different genes might play a potential role in the pathogenesis of these diseases.

Received: 25.09.2019 **Accepted:** 29.11.2019 **Available online:** 10.01.2020
Corresponding Author: Elif Pala, SANKO University, Faculty of Medicine, Department of Medical Biology, Gaziantep, Turkey
E-mail: epala@sanko.edu.tr

Obtaining the affected brain tissue to search the underlying pathology of neurological disorders is the major difficulty for researches. In such cases, the investigation of gene expressions from peripheral blood is a useful method that can identify biomarkers for disease progression and contribute to its pathogenesis. In the literature, there are studies using blood samples to determine potential pathways and genes associated with trinucleotide repeat disorders such as Huntington's disease (15) and Friedreich's ataxia (16).

In our study, gene expression profiles in the Gene Expression Omnibus (GEO) database were analyzed by the Principal Component Analysis (PCA)-based unsupervised Feature Extraction (FE) method to identify differentially expressed genes between FMR1 premutation carriers and healthy controls (normal number of repeats). Identification of potential biomarker genes by using this method will contribute to the investigation of the disease in particular in terms of other related disorders, the determination of the risk of having an affected child and the research of the therapeutic targets of the disease.

MATERIAL and METHODS

We searched the GEO database (https://www.ncbi.nlm.nih.gov/geo/) by using the keywords: "Fragile X" (study keyword), "Homo sapiens" (organism), "Expression profiling by array" (study type). The inclusion criteria were selected as; (1) peripheral blood samples of Fragile X premutation carrier patients compared with healthy controls, (2) adequate data to perform the analysis. The GSE48873 gene expression profile was downloaded from the GEO database, which consists of nine Fragile X premutation male carriers and five healthy male controls. In the relevant study, gene expression profiling had been performed using the Agilent microarray (SurePrint G3 Human GE8×60K Microarray).

PCA is a mathematical data reduction method and the process of extracting relevant information from a large dataset. In contrast to standard PCA, which integrates the samples, PCA-based unsupervised FE integrates the genes (17). It has been previously applied to gene expression data obtained from microarray experiments (17,18). We performed PCA-based unsupervised FE method to reduce the number of predictor variables (genes) furthermore for classification of Fragile X premutation carriers and healthy control samples. Firstly, the GSE48873 gene expression profile series matrix was downloaded for the statistical analysis. "prcomp" R code was performed to obtain principal component (PC) loadings and "lm" R code was used to determine p-values for each PC loadings that $p<0.05$ were selected as significant. Secondly, based on the significant PC scores, "pchisq" R code was applied to calculate the p-values for each gene. p-values were adjusted by "p.adjust" R code and significant genes were selected ($p<0.001$). Then, "prcomp" R code was practiced on the expression profile matrix of the significant genes to get the PC loadings, and "lm" R code was used to calculate p-values ($p<0.05$). Finally, the "lda" (Linear Discriminate Analysis) R code was applied to the PC loadings to classify the samples into two categories (Fragile X premutation carriers/healthy controls). The cross-validation was achieved via "leave-one-out" technique to prevent overfitting (19). Sensitivity, specificity, and Area Under the ROC Curve (AUC) for the optimal cut-point were determined based on discriminant function scores that were obtained by LDA (20). All statistical analyses were done by using R Studio software program.

RESULTS

In our study, the entire data matrix of GSE48873 gene expression profile was [14x62976]. By applying the "prcomp" and "lm" R codes to this matrix, one PC loading (PC2) was selected with an adjusted p-value<0.05.

Table 1. Determined 14 gene/lncRNA names by using PCA-based unsupervised FE method

No	Gene Symbol	Gene Name
1	lincRNA: chr10:17250419-17261819_F	lincRNA: chr10:17250419–17261819 forward strand
2	lincRNA: chr17:15385375-15392270_F	lincRNA: chr17:15385375-15392270 forward strand
3	lincRNA: chr4:68294580-68337980_F	lincRNA: chr4:68294580-68337980 forward strand
4	lincRNA: chr1:86062087-86107987_R	lincRNA: chr1:86062087-86107987 reverse strand
5	ARFRP1	ADP-ribosylation factor related protein 1
6	lincRNA: chr1:181063077-181073127_R	lincRNA: chr1:181063077-181073127 forward strand
7	ENST00000381654	Not found in "Ensembl"
8	LOC100129931	uncharacterized LOC10012993
9	lincRNA: chr22:46451236-46516536_R	lincRNA: chr22:46451236-46516536 reverse strand
10	lincRNA: chr19:53158148-53193727_R	lincRNA: chr19:53158148-53193727 forward strand
11	lincRNA: chr17:73598283-73599500_F	lincRNA: chr17:73598283-73599500 forward strand
12	LOC100133286	uncharacterized LOC100133286
13	WASH5P	WAS protein family homolog 5 pseudogene (WASH5P), lncRNA
14	lincRNA: chr19:52588079-52597393_F	lincRNA: chr19:52588079-52597393 forward strand

After that based on this PC scores, 14 genes were detected as significant with an adjusted p-value<0.001 by using "pchisq" R code. It was determined that the majority of these 14 genes, which showed different expression levels between the groups, consisted of lncRNA (long non-coding RNA), but one of them was not found in the "Ensembl" database (Table 1).

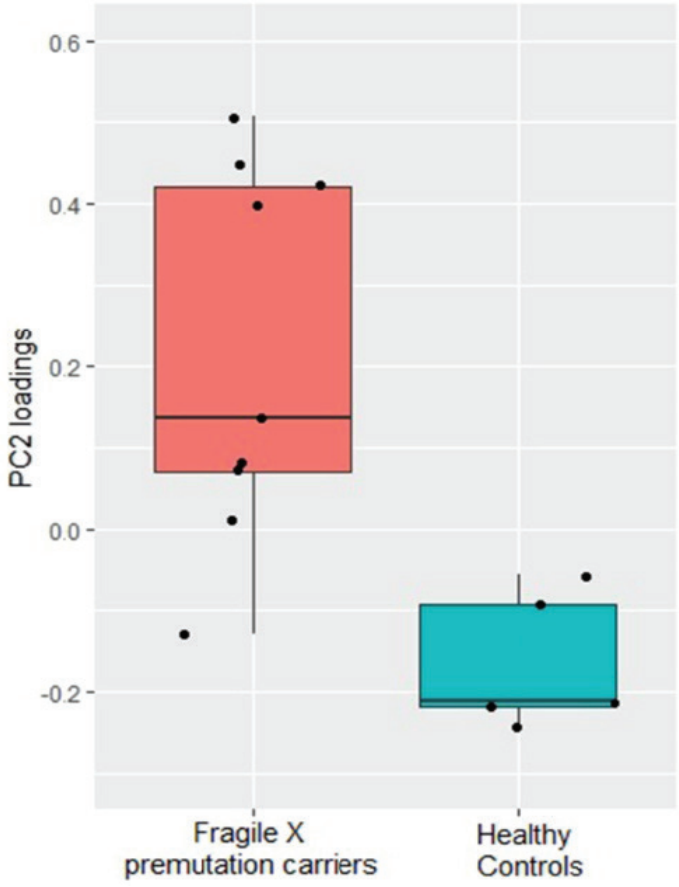

Figure 1. PC2 loading box plot of healthy controls and Fragile X premutation carriers

Once again PC2 was found significant with an adjusted p-value<0.05 when "prcomp" and "lm" R codes were applied to these 14 genes expression profile matrix (Figure 1). Then 14 samples (Fragile X premutation carriers and healthy controls) were discriminated into two classes using the "lda" R code. After that, it was determined that LDA could successfully discriminate these groups (p=0.02098). In a group of five healthy controls, all were classified as healthy, and in a group of nine Fragile X premutation carriers, only one case was misclassified as healthy control (Table 2). ROC analysis with a discriminant function score produced AUC of 0.89 for discriminating Fragile X premutation carriers from healthy control samples with sensitivity and specificity values 89% and 100%, respectively (Figure 2). Also, the overall accuracy was found 80%. Analysis of ROC curve showed that the set of these 14 genes/lncRNAs could be utilized as a potential biomarker for Fragile X premutation carriers.

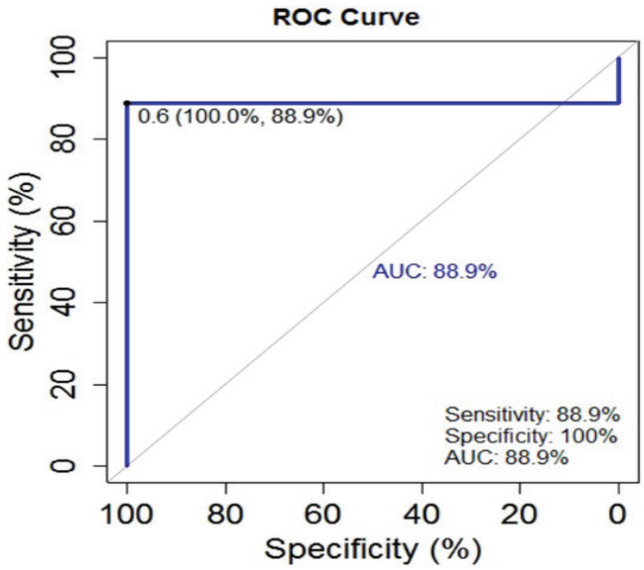

Figure 2. ROC curve obtained by comparing Fragile X premutation carriers with healthy controls

Table 2. Discrimination of patients by using LDA

	Healthy Controls	Fragile X premutation carriers
Healthy Controls	5 (TN)	1 (FN)
Fragile X premutation carriers	-	8 (TP)
Total	**5**	**9**

Abbreviations: TN; true negative, TP; true positive, FN; false negative

DISCUSSION

Trinucleotide repeat disorders are known to cause more than thirty neurological and neuromuscular diseases such as Huntington's disease, FXS, and Spinocerebellar Ataxia (21,22). FXS is an X-linked genetic disorder characterized by hereditary mental retardation, behavioral problems, and specific physical dysmorphisms. It is caused by the increase in CGG repeats in the 5'UTR region of the FMR1 gene.

The FMR protein encoded by the FMR1 gene is an RNA-binding protein and is thought to cause FXS-related mental retardation due to its functions such as translational repression, synaptic maturation, dendritic mRNA localization (23-26).

Approximately 46% of men and 17% of women are at risk for the development of FXTAS in FMR1 premutation carriage (27). FXTAS is characterized by parkinsonism, ataxia, tremor, cognitive decline, and psychiatric symptoms that generally begin later age 50 (8).

In this study, the gene expression profiles of FX premutation carriers and healthy controls were compared by using PCA-based unsupervised FE method. As a result of the analysis, it was found that the set of 14 genes/lncRNAs could be able to differentiate groups with high sensitivity, specificity, and accuracy.

lncRNAs are a group of transcripts of more than 200 nucleotides in length without protein-coding potential (28). Depending on their genomic localization and proximity to the protein-encoding gene, lncRNAs are divided into various classes. In our study, the predominantly identified lincRNAs (long intergenic ncRNA) are types of transcribed within known protein-coding genes (29). It has been reported that lncRNAs show a variety of expression in the central nervous system cells and 5458 of a total of 9747 lncRNA transcripts identified in the human brain are highly expressed in these cells (30). These lncRNAs have been shown to be involved in several critical stages of brain development and function, like synaptogenesis, neurogenesis, and GABAergic interneuron function. Abnormalities in these stages have been reported to be effective in various neurodevelopmental and neurodegenerative diseases (31). In addition, lncRNAs emerge as major regulators of neurogenesis. In studies of loss of function in embryonic stem cells and induced pluripotent stem cells, the differentiation pathways have been shown to be defective as a result of silencing by lncRNA, and numerous lncRNAs have been defined as integral components in neurogenesis (32,33).

Some researchers showed that FMR protein functions as a translational suppressor by binding to target mRNAs to regulate protein synthesis, which is essential for neuronal development (34-36). It is also known that FMR protein is genetically and biochemically involved in the microRNA pathway and that FMR protein-associated miRNAs regulate neuronal development. Consequently, the interaction between FMR protein and non-coding RNA is known to contribute to the pathogenesis of FXS (37).

Studies have shown that lncRNAs play an important role in the pathogenesis of both premutation and full mutation carriers of FMR1 and are particularly effective in nervous system disorders. Differentially expression of lncRNAs in this disease and origination of many related lncRNAs from the FMR1 gene locus in both premutation carriers and FXS patients has also been reported (38,39). This proves that lncRNAs can be used as biomarkers in the diagnosis or assessment of this disease.

CONCLUSION

In this study, genes/lncRNAs were identified using bioinformatics techniques that discriminate Fragile X premutation carriers and healthy controls with high accuracy. These results suggest that lncRNAs may be effective in the pathogenesis of FXTAS and have potential to be used as therapeutic targets which should be validated with further experimental studies.

Financial Disclosure: There are no financial supports.

Ethical approval: This study was approved by the Institutional Ethics Committee and conducted in compliance with the ethical principles according to the Declaration of Helsinki.

Elif Pala ORCID: 0000-0002-1690-3170

REFERENCES

1. Orr HT, Zoghbi HY. Trinucleotide repeat disorders. Annu Rev Neurosci 2007;30:575-621.
2. Verkerk AJ, Pieretti M, Sutcliffe JS, et al. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 1991;65:905-14.
3. Saldarriaga W, Tassone F, González-Teshima LY, et al. Fragile X syndrome. Colomb Med 2014;45:190-8.
4. Fu Y-H, Kuhl DP, Pizzuti A, et al. Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. Cell 1991;67:1047-58.
5. Peprah E. Fragile X syndrome: the FMR1 CGG repeat distribution among world populations. Ann Hum Genet 2012;76:178-91.
6. Tassone F, Hagerman RJ, Garcia-Arocena D, et al. Intranuclear inclusions in neural cells with premutation alleles in fragile X associated tremor/ataxia syndrome. J Med Genet 2004;41:43.
7. Hessl D, Tassone F, Loesch DZ, et al. Abnormal elevation of FMR1 mRNA is associated with psychological symptoms in individuals with the fragile X premutation. Am J Med Genet B Neuropsychiatr Genet 2005;139:115-21.
8. Hagerman RJ, Leehey M, Heinrichs W, et al. Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. Neurology 2001;57:127-30.
9. Sullivan A, Marcus M, Epstein M, et al. Association of FMR1 repeat size with ovarian dysfunction. Hum Reprod 2005;20:402-12.
10. Coffey SM, Cook K, Tartaglia N, et al. Expanded clinical phenotype of women with the FMR1 premutation. Am J Med Genet A 2008;146:1009-16.
11. Hagerman RJ, Leavitt B, Farzin F, et al. Fragile-X–associated tremor/ataxia syndrome (FXTAS) in females with the FMR1 premutation. Am J Hum Genet 2004;74:1051-6.
12. Chiurazzi P, de Graaff E, Ng J, et al. No apparent involvement of the FMR1 gene in five patients with phenotypic manifestations of the fragile X syndrome. Am J Hum Genet 1994;51:309-14.
13. Reyniers E, Wolff G, Tariverdian G, et al. Severe mental retardation and macroorchidism without mutation in the FMR1 gene. Am J Hum Genet 1996;64:408-12.
14. Jacquemont S, Hagerman RJ, Leehey MA, et al. Penetrance of the fragile X–associated tremor/ataxia syndrome in a premutation carrier population. Jama 2004;291:460-9.
15. Borovecki F, Lovrecic L, Zhou J, et al. Genome-wide expression profiling of human blood reveals biomarkers for Huntington's disease. Proc Natl Acad Sci USA 2005;102:11023-8.
16. Haugen AC, Di Prospero NA, Parker JS, et al. K. Altered gene expression and DNA damage in peripheral blood cells from Friedreich's ataxia patients: cellular model of pathology. PLoS Genet 2010;6:1000812.

17. Taguchi Y. Identification of more feasible microRNA–mRNA interactions within multiple cancers using principal component analysis based unsupervised feature extraction. Int J Mol Sci 2016;17:696.
18. Taguchi Y, Wang H. Genetic association between amyotrophic lateral sclerosis and cancer. Genes 2017;8:243.
19. Brenning A, Lausen B. Estimating error rates in the classification of paired organs. Stat Med 2008;27:4515-31.
20. Fawcett T. An introduction to ROC analysis. Pattern Recogn Lett 2006;27:861-74.
21. Castel AL, Cleary JD, Pearson CE. Repeat instability as the basis for human diseases and as a potential target for therapy. Nat Rev Mol Cell Biol 2010;11:165.
22. Mirkin SM. Expandable DNA repeats and human disease. Nature 2007;447:932.
23. 23. Antar L, Dictenberg J, Plociniak M, et al. Localization of FMRP-associated mRNA granules and requirement of microtubules for activity-dependent trafficking in hippocampal neurons. Genes, Brain Behav 2005;4:350-9.
24. Weiler IJ, Irwin SA, Klintsova AY, et al. Fragile X mental retardation protein is translated near synapses in response to neurotransmitter activation. Proc Nat Acad Sci USA 1997;94:5395-400.
25. Irwin SA, Galvez R, Greenough WT. Dendritic spine structural anomalies in fragile-X mental retardation syndrome. Cereb Cortex 2000;10:1038-44.
26. O'Donnell WT, Warren ST. A decade of molecular studies of fragile X syndrome. Annu Rev Neurosci 2002;25:315-38.
27. Garcia-Arocena D, Hagerman PJ. Advances in understanding the molecular basis of FXTAS. Hum Mol Genet 2010;19:83-9.
28. Iyer MK, Niknafs YS, Malik R, et al. The landscape of long noncoding RNAs in the human transcriptome. Nat Genet 2015;47:199.
29. Li X, Wu Z, Fu X, et al. lncRNAs: insights into their function and mechanics in underlying disorders. Mutat Res Rev Mutat Res 2014;762:1-21.
30. Knauss JL, Sun T. Regulatory mechanisms of long noncoding RNAs in vertebrate central nervous system development and function. Neuroscience 2013;235:200-14.
31. Tang J, Yu Y, Yang W. Long noncoding RNA and its contribution to autism spectrum disorders. CNS Neurosci Ther 2017;23:645-56.
32. Ng SY, Johnson R, Stanton LW. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. EMBO J 2012;31:522-33.
33. Lin M, Pedrosa E, Shah A, et al. RNA-Seq of human neurons derived from iPS cells reveals candidate long non-coding RNAs involved in neurogenesis and neuropsychiatric disorders. PLoS One 2011;6:23356.
34. Darnell JC, Klann E. The translation of translational control by FMRP: therapeutic targets for FXS. Nat Neurosci 2013;16:1530.
35. Guo W, Ceolin L, Collins KA, et al. Elevated CaMKIIα and hyperphosphorylation of Homer mediate circuit dysfunction in a fragile X syndrome mouse model. Cell Rep 2015;13:2297-311.
36. Zhang YQ, Bailey AM, Matthies HJ, et al. Drosophila fragile X-related gene regulates the MAP1B homolog Futsch to control synaptic structure and function. Cell 2001;107:591-603.
37. Jin P, Zarnescu DC, Ceman S, et al. Biochemical and genetic interaction between the fragile X mental retardation protein and the microRNA pathway. Nat Neurosci 2004;7:113.
38. Ladd PD, Smith LE, Rabaia NA, et al. An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals. Hum Mol Genet 2007;16:3174-87.
39. Khalil AM, Faghihi MA, Modarresi F, et al. A novel RNA transcript with antiapoptotic function is silenced in fragile X syndrome. PLoS One 2008;3:1486.