Ferdowsi University of Mashhad

ISSN 2008-9147
Numbers: 23

JCMR

Journal of Cell and Molecular Research

Volume 12, Number 1, Summer, 2020

بسم الله الرحمن الرحيم

Issuance License No. 124/902-27.05.2008 from Ministry of Culture and Islamic Guidance
Scientific Research Issuance License No. 161675 from the Ministry of Science, Research and Technology, Iran

Journal of Cell and Molecular Research (JCMR)

Volume 12, Number 1, Summer, 2020

JCMR Office: Department of Biology, Faculty of Sciences, Ferdowsi University of Mashhad, Mashhad, Iran.
Postal Code: 9177948953
P.O. Box: 917751436
Tel: +98-513-8804063/+98-9156996326
Fax: +98-513-8795162
E-mail: jcmr@um.ac.ir
Online Submission: http://jcmr.um.ac.ir

Table of Contents

Journal of Cell and Molecular Research (2020) 12 (1), 1-9 DOI:10.22067/jcmr.v12i1.85654

Research Article

Network Analysis of Differential Gene Expression to Identify Hub Genes in Ovarian Cancer

Akram Siavoshi[1], Mahdieh Taghizadeh[2], Elahe Dookhe[3], Mehran Piran[4], Mahsa Saliani[5], Shahla Mohammad Ganji[6*]

[1] *Department of Biology, Sinabioelixir Group, Alborz Health Technology Development Center, Tehran, Iran*
[2] *Department of Medical Genetic, Tarbiat Modares University, Tehran, Iran*
[3] *Department of Biology, Research and Science Branch, Islamic Azad University, Tehran, Iran*
[4] *Department of Medical Biotechnology, Drug Design and Bioinformatics Unit, Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran*
[5] *Department of Chemistry, Faculty of Science, Ferdowsi University of Mashhad, Mashhad, Iran*
[6] *Institute of Medical Biotechnology, National Institute of Genetic Engineering and Biotechnology, Tehran, Iran*

Received 17 February 2020 Accepted 2 May 2020

Abstract

Epithelial ovarian cancer (EOC), as a challenging disease among women with poor prognosis and unclear molecular pathogenesis, each year is responsible for 140000 deaths globally. Recent progress in the field revealed the importance of proteins as key players of different biological events. Considering the complicated protein interactions, taking a deeper look at protein-protein interactions (PPIs) could be considered as a superior strategy to unravel complex mechanisms encountered with regulatory cell signaling pathways of ovarian cancer. Hence, PPI network analysis was performed on differentially expressed genes (DEGs) of ovarian cancer to discover hub genes which have the potential to be introduced as biomarkers with clinical utility. A PPI network with 600 DEGs was constructed. Network topology analysis determined UBC, FN1, SPP1, ACTB, GAPDH, JUN, and RPL13A, with the highest Degree (K) and betweenness centrality (BC), as shortcuts of the network. KEGG pathway analysis showed that these genes are commonly enriched in ribosome and ECM-receptor interaction pathways. These pivotal hub genes, mainly UBC, FN1, RPL13A, SPP1, and JUN have been reported previously as potential prognostic biomarkers of different types of cancer. However, further experimental molecular studies and computational processes are required to confirm the function and association of the identified hub genes with epithelial ovarian cancer prognosis.

***Keywords*:** Epithelial Ovarian Cancer, Differentially Expressed Gene Analysis, PPI Network Analysis, Pathway Enrichment Analysis

Introduction

Epithelial ovarian cancer (EOC) as a challenging disease is diagnosed in nearly a quarter of a million women, and it is responsible for 140000 deaths worldwide per annum (Krzystyniak et al., 2016; Torre et al., 2018). Lack of early diagnosis and empirically-validated treatments were considered as the most common causes of mortality (Cho et al., 2015). During the past decade, extensive research has been conducted to identify methods to predict and evaluate cancer progression (Li et al., 2015; Loghmani et al., 2014). Currently, the use of biomarkers such as serum cancer antigen 125 (CA125) and human epididymis protein 4 (HE4) is very common among all methods used to diagnose, prognose, and management of ovarian cancer (Archana et al., 2013). However, the potential of these biomarkers for efficient prediction of outcome remains a significant challenge regardless of the different stages and complexity of the disease. Thus, the survival rate of EOC is still low, and there has not been any remarkable success in treatment, especially in patients with advanced epithelial ovarian cancer.

Recently, some researchers proposed some target genes with specific coverage of a determined stage of the disease (Arnedos et al., 2019; Li et al., 2018; Zhang et al., 2019). In this regard, differentially expressed gene analysis (DEGA) (Anders et al., 2010) as the most important application of RNA-Seq experiments, can be used to compare different genes expression levels between normal and cancerous cells. The results of such analyses reveal a list of differentially expressed genes (DEGs). Although, as most human cancers are very complex, and are

*Corresponding author's e-mail address: *Shahla@nigeb.ac.ir*

encountered with sets of genes and their complicated interactions, identification of the exact molecular mechanism is very difficult, especially, in human cases.

Investigating protein-protein interactions (PPIs), with key roles in the biological function of the cells, is one of the beneficial methods to discover complex molecular mechanisms which are responsible for cell signaling and cell to cell communications (Huang et al., 2016). In a previous study, PPI networks were created based on the DEGs analysis to discover hub genes that have the potential to be introduced as biomarkers of esophageal squamous cell carcinoma (Wu et al., 2014).

In the present study, RNA-seq data obtained from normal and cancerous cells of the ovarian tissue were compared statistically to discover DEGs. Then, Systems biology analyses such as gene ontology (GO) and pathways analysis (KEGG) (Ashburner et al., 2000; Kanehisa et al., 2000) were performed to provide insights into the key cellular processes which are responsible for normal/diseased condition. Furthermore, to explore new biomarkers, the PPI network was constructed by mapping all determined DEGs to the network data. Functional enrichment analysis was performed to assign functional categories to the subnetworks of genes.

By introducing several hub genes, results of the present study may facilitate our vision regarding the molecular mechanisms involved in ovarian cancer pathogenesis. These experiments could be effective for defining proper treatment strategies in the clinical settings. However, further confirmatory studies are required for validation of data and announcement of novel panels of genes.

Materials and Methods

RNA sequencing data processing and differential gene expression analysis

Three separate Fastq data files for untreated ovarian tumor cell line SKOV-3, and normal cell line FT194 (De Cristofaro et al., 2016) were retrieved from the sequence read archive (SRA; http://www.ncbi.nlm.nih.gov/geo/). In order to provide clean data for downstream analyses, quality filtration was conducted to omit low quality sequence reads (more than 30% of reads) and adaptors (the first 15bp of Illumina reads) by the Trimmomatic program (Bolger et al., 2014). Then, using the HISAT2 alignment program (Kim et al., 2015) all clean reads were mapped to the Homo sapiens (human) genome assembly GRCh37 (hg19). Counting of transcripts (mapping efficiencies (95%)) was performed with HTSeq (Anders et al., 2014). Count data normalization was performed to determine statistically significant DEGs across two conditions. The DESeq2-Bioconductor package (version 1.6.3) was applied to improve stability and interpretability of estimates. Adjusted P value<0.01 and a |log2FC|>2 were defined as the cut-off criteria. Biological significance of DEGs was explored through GO term enrichment analysis including biological processes (BP), cellular components (CC), and molecular functions (MF), and then KEGG pathway enrichment analysis was performed using enricheR-Bioconductor package (version 2.1).

PPI Network construction

The search tool for the retrieval of interacting genes (STRING, https://string-db.org/cgi/input.pl; version: 11.0), was used for obtaining direct (physical) and indirect (functional) human PPI networks (PPIN). The attribute that we applied to construct network was based on the highest confidence score of 0.07. Then, the constructed PPI network was analyzed using Cytoscape (version 3.7.0). The topological analysis of the PPI network was performed with the Network Analyzer. Betweenness centrality (BC), closeness centrality (CC), and degree (K) were considered as fundamental parameters during our experiments to determine node properties.

Identification of modules and functional annotation analysis

The Molecular complex detection (MCODE) plugin was applied to visualize the significant gene modules in EOC with default parameters and the maximum depth of 100. Selection criteria for top 3 significant modules were set as follows: MCODE scores ≥ 6, and number of nodes ≥ 10. Functional enrichment analysis for each module was performed using g:Profiler (URL: http://biit.cs.ut.ee/gprofiler/).

Results

Differential gene expression analysis results

Distribution of expression values across samples before and after normalization was evaluated to ensure that expression values were similar across normalized counts.

Complete plot of raw counts using log2 transformation (log2 (Non-normalized counts+1)) and then a plot of normalized counts using the DESeq2 are shown in Figure 1 (A) and (B), respectively.

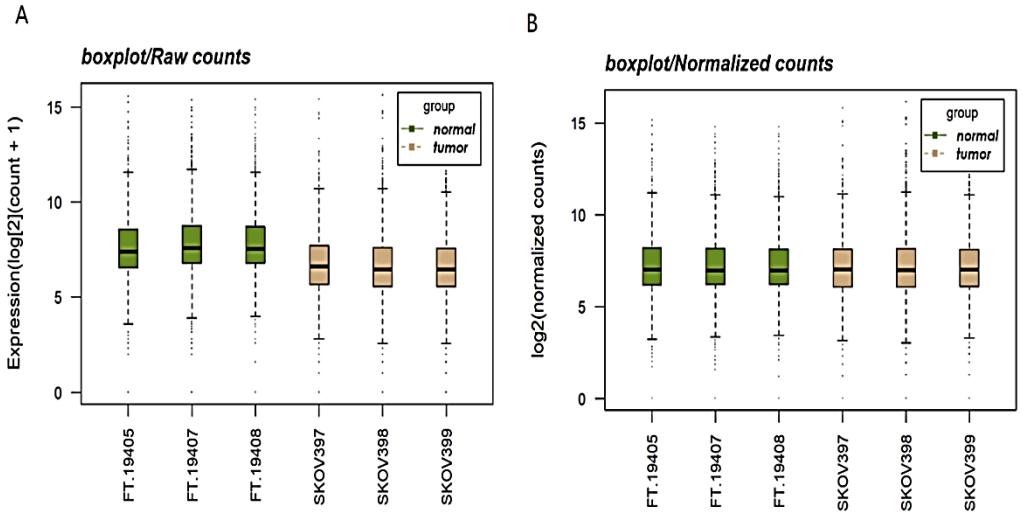

Figure 1. The gene count distributions. Box plots of non-normalized counts (log2 (counts+1)) per sample (A), and normalized counts (log2 (normalized counts)) per sample (B) are shown. The x-axis represents samples and the Y-axis represents log2 (counts +1).

After DE analysis between ovarian tumor and normal groups, 1000 DEGs (padj <0.01) were obtained with 232 upregulated genes (padj <0.01, log2 FC> 2), and 324 downregulated genes (padj <0.01, log2 FC <-2) (Table S1).

Functional analysis of DEGs

Gene ontology (GO) and pathway analysis of differentially expressed genes

To investigate activated and suppressed DEGs in different functional categories, GO and KEGG pathway analysis using enrichR were performed. Based on the results from these experiments genes were classified into different functional categories according to the GO term for biological processes (BP) (Figure. S1), molecular functions (MF) (Figure. S2) and cellular components (CC) (Figure. S3). Totally 572 out of 1000 profiled DEGs assigned to 930 GO terms (padj <0.01). The top 1 significantly upregulated and downregulated GO categories are shown in Table. 1.

Overall, 156 upregulated genes (padj<0.01, log2 FC >2), and 82 downregulated genes (padj<0.01, log2 FC <-2) were mapped to 283 KEGG pathways. The top 14 enriched pathways are shown in Figure. 2. The upregulated genes were highly clustered in signaling pathways including glycolysis, pyruvate metabolism, tryptophan metabolism, and fatty acid degradation; while, the most downregulated genes were highly clustered into ribosome, salmonella infection, focal adhesion, and apoptosis.

PPI Network construction

After DEGA, the significant result of String analyses was based on confidence score (0.007), the average degree of nodes (5.11), and average local clustering coefficient (0.406), and a PPI network with 797 interactions between 600 DEGs was performed (Figure. 3). In order to detect the key parameters of the network, interaction pairs of the PPI network were visualized by Network-Analyzer Cytoscape plugin (cut off values: BC> 0.02, and K> 10)(Table. 2).

Identification of modules and functional annotation analysis

The module analysis of PPI network using MCODE resulted in 13 modules. According to the Table S2 and Figure 4, four significant modules were identified with MCODE (score ≥ 5 and nodes ≥ 6). Among which UBC (Ubiquitin C) as the main hub was clustered in module 3. Ribosomal Protein Small (RPS) subunit genes and Ribosomal Protein Large (RPL) genes were clustered in module 1, and other hub genes including Secreted Phosphoprotein 1 (SPP1), calumenin (CALU), complement C3 (C3), and Fibronectin 1 (FN1) were clustered in module 2.

Table 1. The top 1 enriched gene ontology term of up- and down regulated genes involved in biological processes (BP), cellular components (CC), and molecular functions (MF)

Gene Ontology (GO) terms and ID	Source	Adjusted p-value	Gene symbol
Up-regulated genes			
drug transport (GO:0015893)	BP	0.0197973	SLC47A2; SLC19A1
solute:sodium symporter activity (GO:0015370)	MF	0.06705	SLC5A9; SLC25A22
intrinsic component of the cytoplasmic side of the plasma membrane (GO:0031235)	CC	0.0109396	MIEN1; SPTB
Down-regulated genes			
SRP-dependent cotranslational protein targeting to membrane (GO:0006614)	BP	1.78E-10	RPL41;RPL3;RPL32;RPL13A;RPS25;RPS19;RPL36; RPL14;RPL13;RPL37;RPL26;RPL29;RPS24;RPL19
RNA binding (GO:0003723)	MF	1.56E-13	RBM25;RPL3;RPL32;HMGB2;PSIP1;YBX1;IFIT3;R PS19;RPL36;HIST1H1D;KIF1C;RPL37;HMGN2;HIS T1H1B;HIST1H1C;CAST;DDX58;ACTN1;DNTTIP2 ;RPL13A;PPHLN1;GNL2;GTF2F1;SMC1A;RANGA P1;EEF1D;MYH9;LUC7L3;RPL26;SREK1;RPL29;E ZR;PLEC;DHX8;SRRT;DDX21;PDCD11;TERT;PES 1;UBC;RPL14;RPL13;FLNA;FLNB;SRSF11;RPL19; RBM39;PRPF38B;RPL41;JUN;KRR1;PRRC2C;DEK; EEF2;RPS25;H1F0;MYBBP1A;ACO1;VIM;CALR;R PS24;WRAP53
cytosolic large ribosomal subunit (GO:0022625)	CC	5.33E-10	RPL41;RPL3;RPL32;RPL36;RPL14;RPL13A;RPL13; RPL37;RPL26;RPL29;RPL19

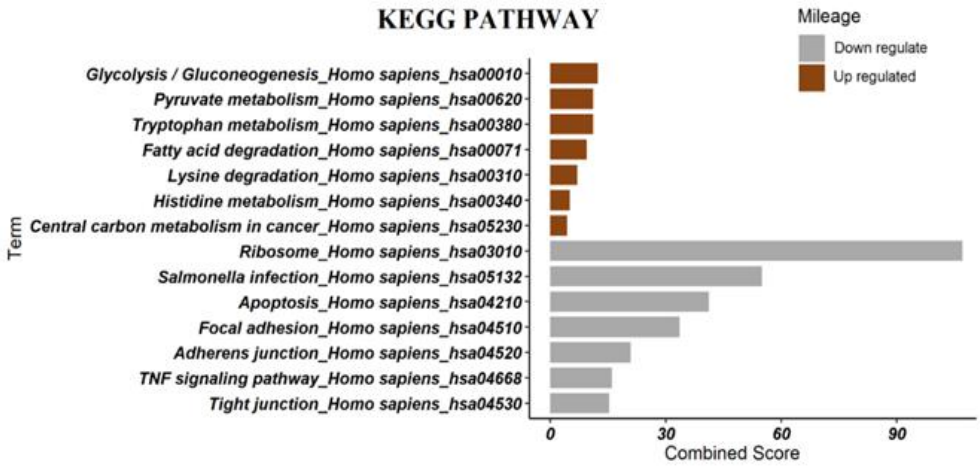

Figure 2. Pathway enrichment analysis of up- and downregulated genes in ovarian cancer samples in comparison to normal cases. The x-axis represents pathways, and KEGG IDs and the Y-axis represents combined score. As shown, the most downregulated genes are enriched in the ribosome pathway and the most upregulated genes are enriched in the Glycolysis/Gluconeogenesis pathway.

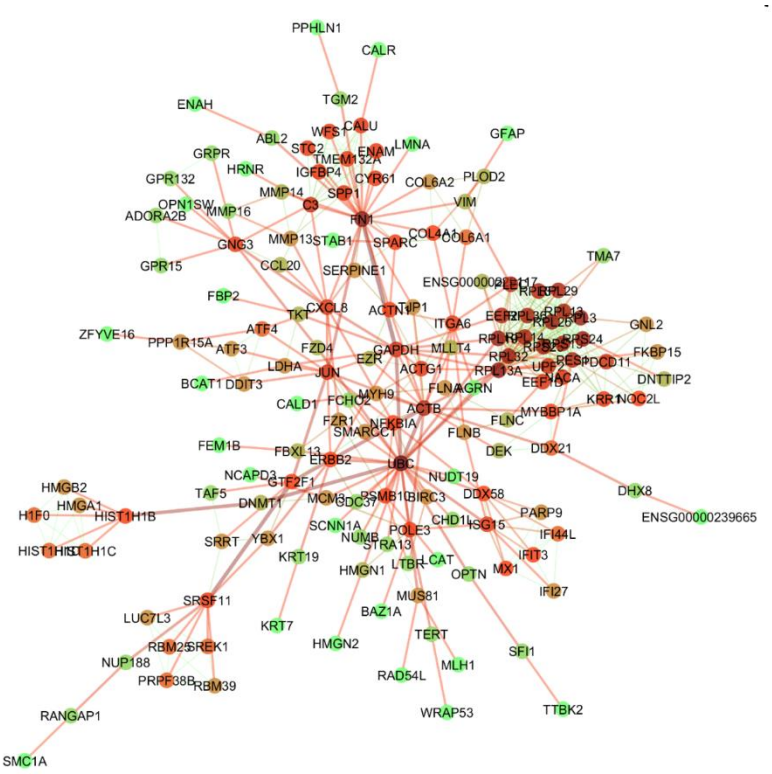

Figure 3. Overview of the PPI network constructed using Cytoscape. The network includes 547 edges (interactions) among 63 nodes. The nodes with dark brown, light brown, and green colors represent key genes in the network. Among key genes, nodes with dark brown color represent the super hubs with the highest BC and K.

Table 2. The main topological parameters including, Betweenness centrality (BC), closeness centrality (CC), and Degree (K) of the PPI network. The hub genes in the network based on cut off values of BC > 0.02 and degree >10 were demonstrated with light gray.

Gene	K	BC	CC
UBC	27	0.39930027	0.40957447
FN1	23	0.22956964	0.35240275
ACTB	17	0.21966548	0.39896373
GAPDH	13	0.19844427	0.40633245
JUN	12	0.12427487	0.36842105
RPL13A	23	0.1023342	0.38118812
CXCL8	11	0.09960491	0.34684685
RPL19	21	0.05464691	0.36150235
RPL32	20	0.04926171	0.35981308
C3	13	0.02902674	0.28308824
PLEC	16	0.02815066	0.30985915
PES1	17	0.02631353	0.29222011

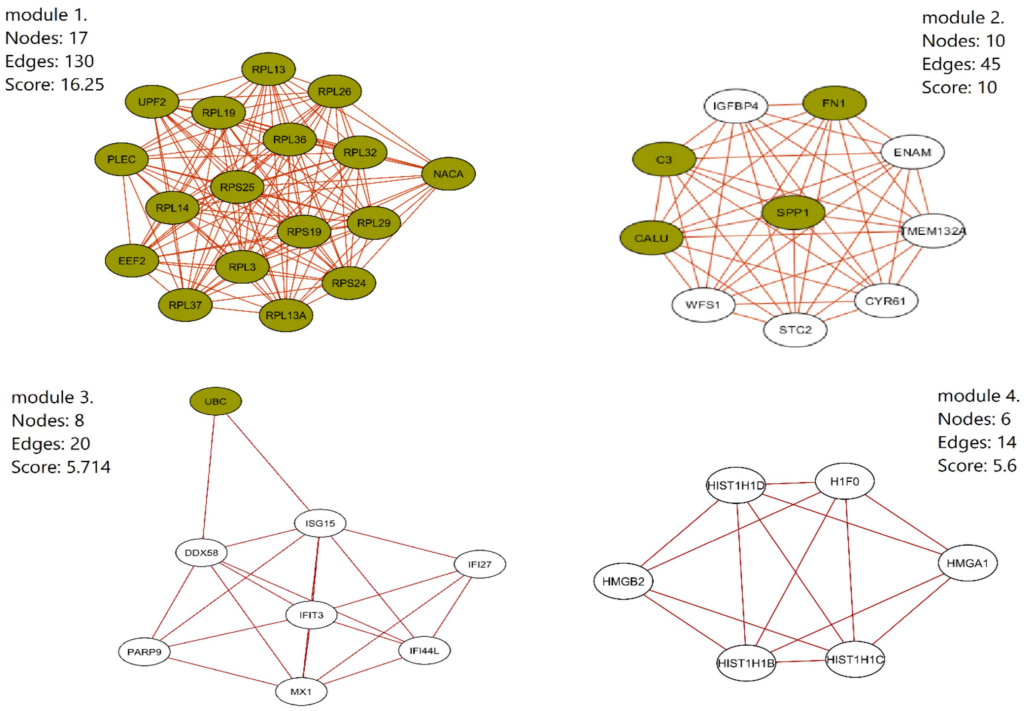

Figure 4. Subnetworks identified from the PPI network (Module 1, Module 2, Module 3, and Module 4, respectively). The light green nodes in each of the clusters represent hub genes which were extracted from the PPI network by Network-Analyzer plugin. The white nodes represent genes which are involved in modules. The lines represent node's interactions.

Functional annotation analysis was applied to the hub genes of the PPI Network and each module separately. The top three functional annotation categories (BB, MF, CC, and KEGG) for module 1 are shown in Table. 3. Pathway analysis mainly involved KEGG pathway revealed that genes were commonly enriched in the ribosome pathway. The constituent structures of the ribosome, nuclear-transcribed mRNA catabolic process, nonsense-mediated mRNA decay, and cytosolic ribosome were the most related terms to MF, BP, and CC of the module 1 with the most enriched gene, respectively.

Table 3. The top three functional annotation categories (BB, MF, CC, and KEGG) for module 1 with the most enriched gene.

Source	GO /Pathway ID	Term name	Adjusted p-value	Intersections
MF	GO:0003735	structural constituent of ribosome	4.98E-19	RPL32,RPS24,RPL26,RPL36,RPS19,RPL3,RPL14,RPL37,RPL19,RPL13A,RPL29,RPL13
MF	GO:0005198	RNA binding	4.26E-13	RPL32,RPS24,RPL26,RPL36,RPS19,RPL3,RPL14,RPL37,RPL19,PLEC,RPL13A,RPL29,RPL13
MF	GO:0003723	structural molecule activity	7.63E-13	RPL32,RPS24,RPL26,RPL36,RPS19,RPL3,RPL14,RPL37,RPL19,RPS25,PLEC,RPL13A,RPL29,UPF2,RPL13,EEF2
BP	GO:0000184	nuclear-transcribed mRNA catabolic process, nonsense-mediated mRNA decay	2.49E-25	RPL32,RPS24,RPL26,RPL36,RPS19,RPL3,RPL14,RPL37,RPL19,RPS25,RPL13A,RPL29,UPF2,RPL13

BP	GO:0006614	SRP-dependent cotranslational protein targeting to the membrane	5.23E-24	RPL32,RPS24,RPL26,RPL36,RPS19,RPL3,RPL14,RPL37,RPL19,RPS25,RPL13A,RPL29,RPL13
BP	GO:0006613	cotranslational protein targeting to the membrane	1.06E-23	RPL32,RPS24,RPL26,RPL36,RPS19,RPL3,RPL14,RPL37,RPL19,RPS25,RPL13A,RPL29,RPL13
CC	GO:0022626	cytosolic ribosome	3.23E-15	NACA,RPL32,RPS24,RPL26,RPL36,RPS19,RPL3,RPL14,RPL37,RPL19,RPS25,RPL13A,RPL29,RPL13,EEF2
CC	GO:0005840	Ribosome	9.54E-14	RPL32,RPS24,RPL26,RPL36,RPS19,RPL3,RPL14,RPL37,RPL19,RPS25,RPL13A,RPL29,RPL13
CC	GO:0044391	ribosomal subunit	7.20E-13	RPL32,RPS24,RPL26,RPL36,RPS19,RPL3,RPL14,RPL37,RPL19,RPS25,RPL13A,RPL29,RPL13,EEF2
KEGG	KEGG:03010	Ribosome	2.85E-19	RPL32,RPS24,RPL26,RPL36,RPS19,RPL3,RPL14,RPL37,RPL19,RPS25,RPL13A,RPL29,RPL13

Discussion

Epithelial ovarian cancer (EOC) has the highest mortality rate among different types of women's cancers due to the poor diagnosis (Hao et al., 2010). Studies have shown that in order to achieve effective methods for early diagnosis and prevention of metastasis, it is important to study the molecular mechanisms of the carcinogenesis.

The purpose of the current study is the analysis of existing RNA-seq data and their comparative interpretation between normal and diseased conditions to investigate novel DEGs involved in PPI Networks and regulatory pathways of EOC. Many pivotal genes and pathways which are associated with ovarian cancer were identified in the present study. Totally, among 1000 DEGs (232 upregulated and 324 downregulated genes), migration and invasion enhancer 1 (MIEN1) and AP001610.5 were the most up- and down-regulated genes, respectively. MIEN1 is an intrinsic component of the cytoplasmic side of the plasma membrane, which plays a pivotal role in the regulation of apoptosis. It was previously proposed by other studies as an important target to be considered in molecular cancer therapy procedures (Evans et al., 2006). The Ribosome pathway and Glycolysis/Gluconeogenesis were also identified as the most significantly enriched pathways in KEGG analysis. ALDH3A2 (Marcato et al, 2011), a member of the aldehyde dehydrogenase (ALDH) gene family (Warburg, 1956), is the most significant gene in Glycolysis/Gluconeogenesis pathway. Whereas, high glycolysis in tumor cells correlates with the degree of tumor malignancy, an argument to justify the significance of the glycolysis pathway in this study is the potential need of chronic cell proliferation to provide energy in order to fuel rapid cell growth and division (Board et al., 1992). In the present study, ribosomal protein (Rps) genes and large ribosomal proteins (RPL) including RPL41, RPL3, RPL32, RPL13A, RPS25, RPS19, RPL14 and RPL36 were the most significant downregulated DEGs. These genes related to the signal-recognition particle (SRP)-dependent cotranslational protein-membrane targeting, RNA binding, and cytosolic large ribosomal subunit.

After analyzing the topology of the PPI network totally 28 nodes with BC> 0.02, and K> 10 were extracted as hub genes and among them some nodes such as UBC, FN1, ACTB, GAPDH, JUN, and RPL13A with high K, BC, and CC were shortlisted. These hub genes were downregulated in EOC samples in comparison to adjacent normal samples. Clustering was performed to investigate the relationship between hub genes with other genes of the network using MCODE and previously identified 13 modules. Furthermore, functional annotation was performed on 4 modules to determine the top affected functions in EOC. Functional annotation of main hubs clustered in these modules showed that Ubiquitin C (UBC), with the highest degree of connectivity, was clustered in module 3 along with other genes, including IFI44L, IFI27, DDX58, IFIT3, PARP9, MX1, and ISG15. This module mainly enriched with protein tag, defense response, and RIG-I-like receptor signaling pathways.

The ubiquitin is encoded by the ubiquitin C (UBC) and ubiquitin B (UBB) in humans. These two genes are essential for maintenance of cellular ubiquitin

levels under stress conditions (Castello et al., 2017). Moreover, they play key roles as tumor suppressors in a variety of cancers, DNA damage repair and regulation of protein turnover through the ubiquitin-proteasome system (UPS) (Kimura et al., 2016). Recent studies indicated that the transcriptional repression of UBB is a cancer-subtype-specific event which occurs in approximately 30% of high-grade serous ovarian cancer (HGSOC) cases. Silencing of UBB reduces cellular ubiquitin levels which is resulted in the overexpression of UBC to compensate the lost function of UBB. These changes may have prognostic value (Dasgupta et al., 2009).

Fibronectin 1 (FN1) was clustered with the Secreted Phosphoprotein 1 (SPP1), Transmembrane Protein 132A (TMEM132A), Stanniocalcin 2 (STC2), cysteine-rich angiogenic inducer 61 (CYR61), Wolfram syndrome type 1 (WFS1), Insulin-like Growth Factor Binding Protein 4 (IGFBP4), and Enamelin (ENAM) in the module 2 which is enriched with extracellular matrix structural constituent. The FN1 has numerous functional properties and is involved in cell adhesion, growth, migration, and differentiation procedures. Previous studies reported morphological alterations in tumors and tumor-derived cell lines that have been attributed to the decrease fibronectin expression, increased fibronectin degradation, and/or decreased expression of fibronectin-binding receptors, such as α5β1 integrin (Zhuo et al., 2016). The main functions of module 1 were correlated with RPs and RPL genes and structural constituent of ribosomes and nuclear-transcribed mRNA catabolic process.

Secreted phosphoprotein 1 (SPP1), also known as Osteopontin (OPN), as an upregulated gene in the present study, was found to be overexpressed in numerous tumors, including lung, colon, breast, and ovarian cancers (Wang et al., 2014; Zeng et al., 2018). Many recent studies demonstrated that the existence of SPP1 in cancerous tissue samples and sera of women with ovarian cancer promotes ovarian cancer progression via Integrin β1/FAK/AKT signaling pathway (Shevde et al., 2014). The SPP1 along with TMEM132A, CALU, C3, STC2, CYR61, WFS1, IGFBP4, FN1, and ENAM were correlated with the most upregulated gene-enriched signaling pathways including post-translational protein modifications, signaling receptor bindings, and ECM-receptor interactions in the module 1. To data, among all mentioned pathways, the ECM-receptor interactions pathway has been highlighted in cancer studies and also the interaction of this pathway with DEGs has been introduced as a diagnostic marker (Bao et al., 2019). The main cancer-related activity of this pathway is related to adhesion, migration, differentiation, proliferation, and apoptosis. Therefore, Increasing the expression of SPP1 as an inflammatory, fibrotic, and carcinogenic gene has been well justified in the ECM-receptor interactions pathway.

Conclusion

The current study demonstrates that, the hub genes derived from the PPI network, including UBC, FN1, ACTB, SPP1, JUN, and RPL13A tend to be present in different cancer-related pathways and Go functions. After following the function of these genes in causing cancer we suggested that these genes may be have potential to become biomarker panel related to the EOC. Yet, more molecular biology experiments, computational method analysis on big data is needed to support this suggestion.

Acknowledgments

Firstly, we would like to thank Dr. Mohammad Ganji for her dedicated support and guidance. And many thanks to all participants who helped to perform this study.

Conflict of interest

The authors declare no conflict of interest.

Reference

Anders S. and Huber W. (2010) Differential expression analysis for sequence count data. Genome Biol, 11:R106.

Anders S., Pyl P. T. and Huber W. (2014) HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics, 31: 166-9.

Archana R., Simmons K. B., Robert C. and Bast J. R. (2013) The Emerging Role of HE4 in the Evaluation of Advanced Epithelial Ovarian and Endometrial Carcinomas. Oncology, 27: 548-556.

Arnedos M., Roulleaux Degage M., Perez-Garcia J. and Cortes J. (2019) Window of Opportunity trials for biomarker discovery in breast cancer. Curr. Opin. Oncol, 31: 486-492.

Ashburner M., Ball C. A., Blake J. A., Botstein D. and Butler H. (2000) Gene Ontology: tool for the unification of biology. Nat. Genet, 25: 25-9.

Bao Y., Wang L., Shi L., Yun F., Liu X., Chen Y., Chen C., Ren Y. and Jia Y. (2019) Transcriptome profiling revealed multiple genes and ECM-receptor interaction pathways that may be associated with breast cancer. Cell Mol Biol Lett, 24(38): 2-20.

Board P. G, Coggan M., Baker R. T., Vuust J. and Webb G. C. (1992) Localization of the human UBC polyubiquitin gene to chromosome band 12q24.3. Genomics, 12: 639-42.

Bolger A. M., Lohse M. and Usadel B. (2014) Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics, 30: 2114-20.

Castello L. M., Raineri D., Salmi L., Clemente N. and Vaschetto R. (2017) Osteopontin at the Crossroads of Inflammation and Tumor Progression. Mediators Inflamm, 22.

Cho A., Howell V. M. and Colvin E. K. (2015) The Extracellular Matrix in Epithelial Ovarian Cancer. A Piece of a Puzzle. Front. Oncol, 5: 245.

Dasgupta S., Wasson L. M., Rauniyar N., Prokai L. and Borejdo J. (2009) Novel gene C17orf37 in 17q12 amplicon promotes migration and invasion of prostate cancer cells. Oncogene, 28: 2860-2872.

De Cristofaro T., Di Palma T., Soriano A., Monticelli A. and Affinito O. (2016) Candidate genes and pathways downstream of PAX8 involved in ovarian high-grade serous carcinoma. J. Oncotarget, 7: 41929-41947.

Evans E. E., Henn A. D., Jonason A., Paris M. J. and Schiffhauer L. M. (2006) C35 (C17orf37) is a novel tumor biomarker abundantly expressed in breast cancer. Mol Cancer Ther, 11: 2919-30S.

Hao J. M, Chen J. Z., Sui H. M., Si-Ma X. Q. and Li GQ. (2010) A five-gene signature as a potential predictor of metastasis and survival in colorectal cancer. J. Pathol, 220: 475-89.

Huang Y. A., You Z. H., Chen X., Chan K. and Luo X. (2016) Sequence-based prediction of protein-protein interactions using weighted sparse representation model combined with global encoding. BMC Bioinformatics, 17: 184.

Kanehisa M. and Goto S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res, 28: 27-30.

Kim D., Langmead B. and Salzberg S. L. (2015) HISAT: a fast spliced aligner with low memory requirements. Nat Methods, 12: 357-60.

Kimura Y. and Tanaka K. (2016) Regulatory mechanisms involved in the control of ubiquitin homeostasis. J. Biochem, 147: 793-8.

Krzystyniak J., Ceppi L., Dizon D. S. and Birrer M. J. (2016) Epithelial ovarian cancer: the molecular genetics of epithelial. Ann. Oncol, 27: i4-i10.

Li M. X., Jin L. T., Wang T. J., Feng Y. J., Pan C. P. and Zhao D. M. (2018) Identification of potential core genes in triple negative breast cancer using bioinformatics analysis. Onco. Targets Ther, 11: 4105-4112.

Li Y., Xiao X., Ji X., Liu B. and Amos C. I. (2015) RNA-seq analysis of lung adenocarcinomas reveals different gene expression profiles between smoking and nonsmoking patients. Tumour. Biol, 36: 8993-9003.

Loghmani H., Noruzinia M., Abdul-Tehrani H., Taghizadeh M. and Karbassiane M. H. (2014) Association of estrogen receptors' promoter methylation and clinicopathological characteristics in Iranian patients with breast cancer. Molecular and Biochemical diagnosis (MBD), 1: 21-33.

Love M. I., Huber W. and Anders S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol, 15: 550.

Marcato P., Dean C. A., Araslanova R., Gillis M. and Joshi M. (2011) Aldehyde dehydrogenase activity of breast cancer stem cells is primarily due to isoform ALDH1A3 and its expression is predictive of metastasis. J. Stem. Cells, 29: 32-45.

Shevde L. A. and Samant R. S. (2014) Role of osteopontin in the pathophysiology of cancer. Matrix Biol, 37: 131–141.

Thor A. D., Young R. H. and Clement P. B. (1991) Pathology of the fallopian tube, broad ligament, peritoneum, and pelvic soft tissues. Hum. Pathol, 22: 856-867.

Torre L. A., Trabert B., DeSantis C. E., Miller K. D. and Samimi G. (2018) Ovarian cancer statistics. CA. Cancer J. Clin, 68: 284-296.

Wang S., Zhou L., Han L. and Yuan Y. (2014) Expression and purification of non-tagged recombinant mouse SPP1 in E. coli and its biological significance. Bioengineered, 5: 405-408.

Warburg O. (1956) On respiratory impairment in cancer cells. Science, 124: 269-70.

Wu B., Xie J., Du Z., Wu J. and Zhang P. (2014) PPI network analysis of mRNA expression profile of ezrin knockdown in esophageal squamous cell carcinoma. Biomed Res Int, 3: 651954.

Zeng B., Min Z., Wu H. and Zhengai X. (2018) SPP1 promotes ovarian cancer progression via Integrin β1/FAK/AKT signaling pathway. Onco Targets Ther, 11: 1333-1343.

Zhang X, Wang Y. (2019) Identification of hub genes and key pathways associated with the progression of gynecological cancer. Oncol Lett, 18: 6516-6524.

Zhuo C., Li X., Zhuang H., Tian S. and Cui H. (2016) Elevated THBS2, COL1A2, and SPP1 expression levels as predictors of gastric cancer prognosis. Cell Physiol Biochem, 40: 1316-1324.

Supplementary Materials:

Supplementary Figures

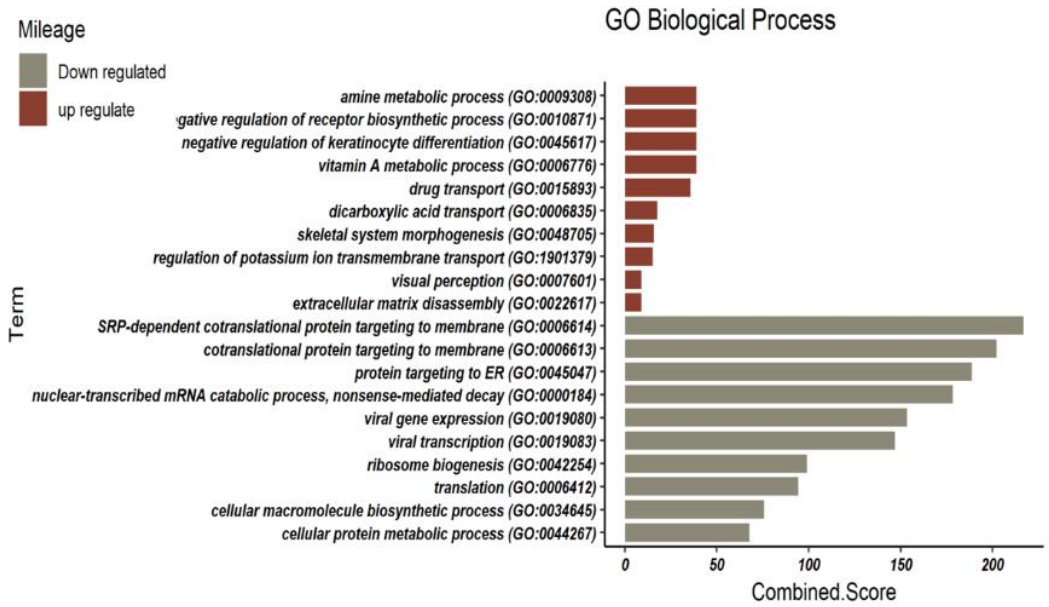

Figure S1. Comparative gene ontology enrichment analysis of biological processes (BP) for up- and down-regulated genes of normal and cancerous ovarian samples. As shown, the most down-regulated genes are enriched in the (SRP)-dependent cotranslational protein-membrane targeting and the most up- regulated genes are enriched in the amino metabolic process.

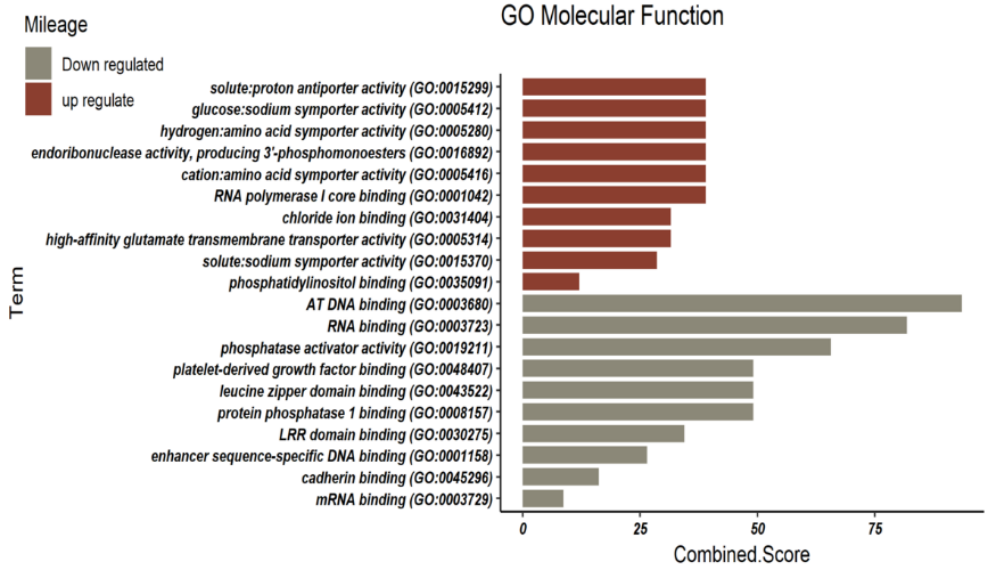

Figure S2. Comparative gene ontology enrichment analysis of molecular functions (MF) for up- and down-regulated genes of normal and cancerous ovarian samples. As shown, the most downregulated genes are enriched in the AT DNA binding and the most upregulated genes are enriched in the Solute: proton antiporter activity.

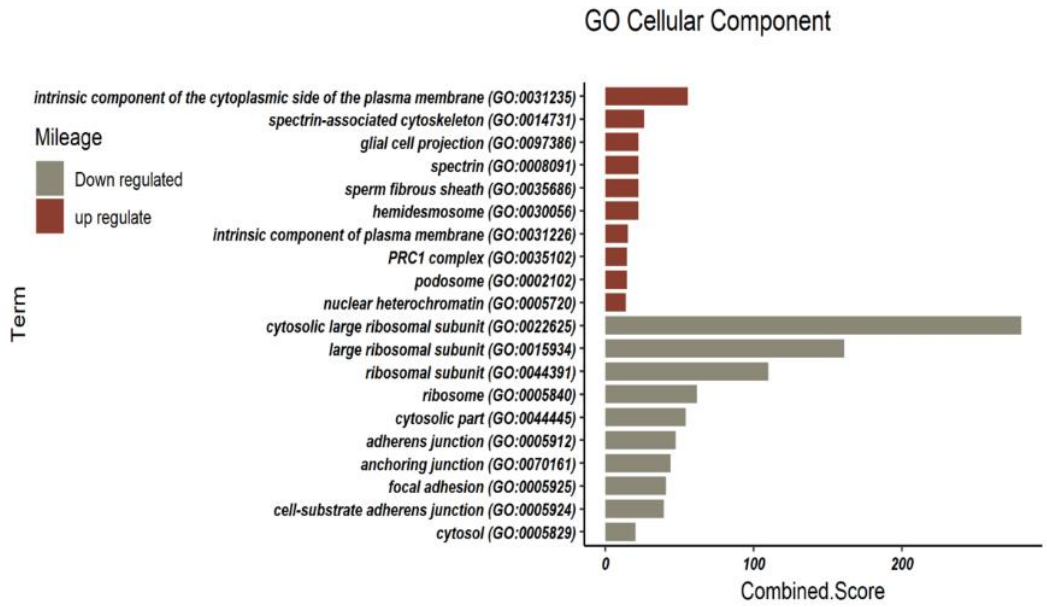

Figure S3. Comparative gene ontology enrichment analysis of cellular components (CC) for up- and down- regulated genes of normal and cancerous ovarian samples. As shown, the most downregulated genes are enriched in cytosolic large ribosomal subunit and the most upregulated genes are enriched in the intrinsic component of the cytoplasmic side of the plasma membrane.

Supplementary Table

Table S1. The list of upregulated and downregulated genes (DEGs). This table is supplied as an excel file.

Research Article

Identification and Functional Profiling of Differentially Expressed Extracellular Vesicle-free MicroRNAs for Efficient Prostate Cancer Diagnosis

Mina Jahandideh, Ebrahim Barzegari*

Medical Biology Research Center, Health Technology Institute, Kermanshah University of Medical Sciences, Kermanshah, Iran

Received 1 June 2020 Accepted 27 July 2020

Abstract

MicroRNAs are interesting as cancer diagnostic and prognostic biomarkers because of their unique tissue expression profiles, higher stability in the blood in comparison to mRNAs, and the possibility for reliable quantification. In the case of prostate cancer (PCa), it is currently emphasized to explore new biomarkers, particularly from microRNAs which are freely available in the bloodstream. In this study, the gene expression omnibus database (GEO), a repository of microarray data for PCa circulating extracellular vesicle-free microRNAs profiling, was analyzed for differentially expressed miRNAs (DE-miRs). Top 20 most differentially expressed miRs with significant (adjusted *p*-value < 0.01) high expression (fold change) levels were extracted by the simultaneous application of different filtering criteria. Then, microRNA-gene networks were constructed for the two sets of positively (n=20) or negatively (n=20) regulated miRNAs. Gene ontology annotations of the target gene sets were also extracted and analyzed. Results indicated that human miR-1587, miR-223-3p, miR-3125, and miR-642b-3p are highly significant DE-miRs in PCa. In addition, human miR-4459, miR-1273g, miR 642a-3p, and miR-642b-3p were identified as top-ranked hubs in the relevant miRNA-gene networks. FOXK1, PML, CD24, ATN1, BAZ2A, CDKN1A, NUFIP2, and HARNPU were identified as microRNA target genes with significant dysregulation. miR-4459, miR-1273g-3p, miR-3135b, miR-5001-5p, and miR-1587 were proposed as novel microRNAs with the potential to be utilized as diagnostic biomarkers of prostate cancer among circulating vesicle-free miRNAs.

***Keywords*:** Prostate Cancer, Diagnosis, Biomarker, Vesicle-free microRNAs, Gene ontology

Introduction

Prostate cancer (PCa), as one of the most prevalent cancer types in males, is estimated to affect more than one million cases annually in the Western world (Bray et al., 2018). Benign forms of PCa can be cured through surgery or radio-therapy with a 5-year survival rate in nearly 100% of cases. However, metastatic cases are not generally treatable (5-year survival rate of less than 40%)(Hamdy et al., 2016; Helgstrand et al., 2018). For patients with advanced stage of the disease, over-treatment is used for long-term patient survival. Accordingly, it is critical to diagnose PCa in a timely and accurate manner.

Transrectal biopsy is the common diagnostic approach of PCa, usually performed following the observation of increased levels of prostate specific antigen (PSA) in the serum, and suspicious results from the low-sensitivity method of digital rectal examination. Although, this ultrasound-guided biopsy technique is erroneous (with false-positive rate 25%)(Bolla and van Poppel, 2012). Multi-parametric magnetic resonance imaging outperforms the standard transrectal biopsy method, but it is accompanied by false negatives and has limited availability (Boesen et al., 2018; Elkjaer et al., 2018). Considering these drawbacks, finding less invasive and more accurate strategies for prostate cancer diagnosis is imperative.

Liquid biopsy samples from blood and urine are convenient sources of possible nucleic acid markers associated with cancer initiation or progression. Circulating non-coding RNAs, specifically microRNAs (miRNAs, miRs), play a significant role in distinguishing various types of tumors, including PCa (Laursen et al., 2019; Lin and Gregory, 2015; Movahedpour et al., 2019; Schmidt et al., 2018; Urabe et al., 2019).

During recent years, miRNAs have brought about a micro-revolution in cancer research and diagnosis due to their contributions as oncogenes or tumor-suppressor genes (Ferracin et al., 2010). There are two forms of circulating cell-free microRNAs within the blood: I) vesicle-enclosed miRNAs (Huang et al., 2013; Valadi et al., 2007), and II) vesicle-free miRNAs (RNA binding proteins) (Wang et al.,

*Corresponding author's e-mail address: *e.barzegari@kums.ac.ir*

2010). As vesicle-free microRNAs are more available for extraction and performing downstream experiments, they could be considered as more appropriate candidates for development of novel detection methods based on the application of miR biomarkers.

Many studies focus on the applications of miRs as biomarkers in PCa detection. Even though, such data have been used fewer for understanding the basic functions of these regulatory elements in normal and cancerous tissues of the prostate. In the current study, we profile regulatory networks and functional targets of vesicle-free circulating microRNAs which are significantly dysregulated in PCa patients. By exploiting the non-coding RNA profiling data, this study aims to unveil the potentials of the new types of microRNAs to be utilized as biomarkers for PCa occurrence or prognosis.

Materials and Methods

Microarray data retrieval and analysis

The miRNA expression profiles of PCa patients vs. healthy individuals were obtained from the national center for biotechnology information (NCBI) gene expression omnibus (GEO) with accession number GSE113234 (Mello-Grand et al., 2019). These data were obtained by microarrays from 60 PCa patients and 27 healthy subjects. The microRNAs were hybridized on arrays with probes designed for a set of 2006 human miRs. During comparative studies (PCa vs. healthy subjects) using the last version of R (version 3.4.1), we retained differentially expressed miRNAs with fold-changes greater than 2.0. *p*-values were adjusted through Benjamini & Hochberg method. Shortlisted circulating microRNAs were then utilized in combination with the results from PSA test to develop an accurate and non-invasive prostate cancer detection strategy.

Identification of differentially expressed microRNAs

To identify the significantly up-regulated microRNAs in patients with PCa in comparison to healthy subjects, miRs with the highest fold change values were extracted from outputs. The same procedure was performed for the lowest values of logFC to identify the significantly down-regulated microRNAs. Different search filters including the existence of statistically significant differences and cut-off values for gene expression fold changes were applied simultaneously to identify top 20 differentially expressed miRs. These DE miRs with the highest or lowest fold changes were selected for further analysis. The *p* values for these sets of miRNAs were less than 0.01 for both positively and negatively regulated microRNAs.

Target identification

MirWalk-database was used to identify the target genes of each DE-miRNA. For each miRNA, a set of the 100 most significantly related genes were determined. Thereby, totally 2000 genes were extracted as the targets of up-regulated and down-regulated miRs.

Functional enrichment analysis

Genes targeted by the most significantly up-regulated microRNAs in PCa were used as primary input to the Database for annotation, visualization and integrated discovery (DAVID, v6.8). This database provides clustering and annotations to the input genes list. Similar procedure was performed for the gene targets of down-regulated microRNAs. These analyses were designed to find enriched biological themes/ gene ontology (GO) terms related to the biological processes (BPs) encountered with the functions of DE-miRNA target genes. Top 20 of the significant GO terms were considered as representatives of the functional annotations for DE-miRNAs with key roles in prostate cancer.

A Visual Basic Script (VBScript) code was created in the Microsoft Excel program to identify the genes which are most frequently present in the collective sets of genes associated with the top BPs. It could be considered that these genes are involved in the pathogenesis or manifestation of PCa.

Network analyses

NetworkAnalyst 3.0 platform was used to extract the complete network of genes which are frequent among the miRs targets (Zhou et al., 2019). Moreover, miRNet 2.0 visualization web-tool was utilized to extract and analyze the miRNA-target gene interaction networks (Chang et al., 2020). The human microRNA disease database (HMDD, v3.2) was screened for evidences confirmed associations between miRs and PCa (Huang et al., 2019). Also, to find driver genes of prostate cancer, DisGeNet (version 7.0), a gene-disease association dataset, was administered. In order to visualized the functional interconnections of the target genes in addition to their significant GO terms, the GOnet server tool was applied using the threshold value p <1.94e-5 as the relation significance (Pomaznoy et al., 2018).

Results

Gene expression value distributions

Cross-platform comparison of gene expression value distributions was performed to determine whether the expression profiles are comparable between PCa and healthy individuals in the corresponding GEO profile entries. As illustrated in Figure 1, both samples demonstrated symmetric distribution of gene expressions. This pattern indicates that the data are median-centered across samples, and cross-group comparison could be applied for further screenings.

MicroRNAs with differential expression levels

Table 1 shows a list of the top 20 miRNAs with significant positively or negatively altered gene expression levels ($p < 0.01$). Based on this analysis, hsa-miR-1587, hsa-miR-223-3p, hsa-miR-1915-3p, hsa-miR-6125, and hsa-miR-1273g-3p are the most significant differentially expressed miRNAs with higher expression levels in patients with prostate cancer vs. healthy controls. On the other hand, hsa-miR-3125, hsa-miR-642b-3p, hsa-miR-4505, hsa-miR-498, and hsa-miR-874 are top 5 down-regulated miRNAs.

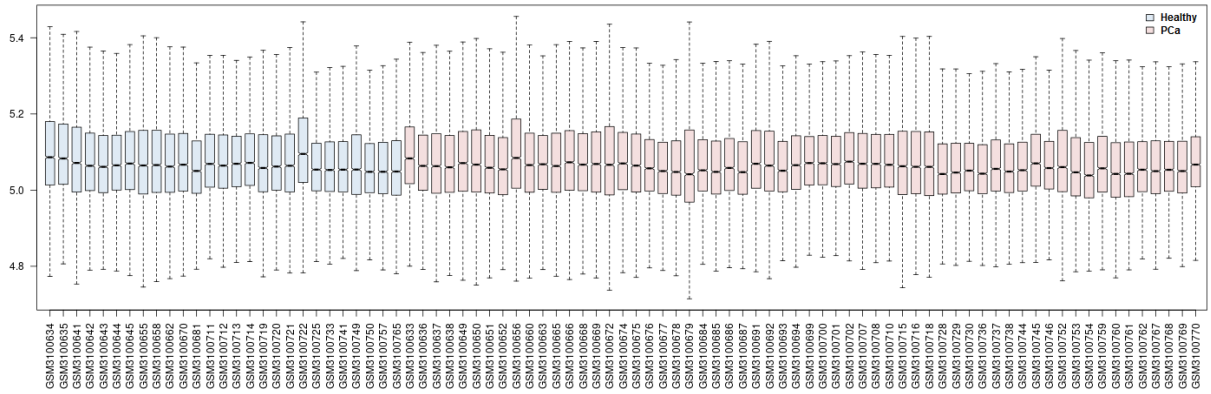

Figure 1. Gene expression value distributions are presented as a grouped boxplot for samples from healthy and PCa patients. GSM identifiers of the samples are shown on the horizontal axis, while, expression values are shown on the vertical axis. Data analysis was performed using built-in distribution analysis tool of the Gene Expression Omnibus database. Abbreviations: GSM, Gene Expression Omnibus Sample entry.

Table 1. Vesicle-free microRNAs differentially expressed in PCa vs. normal subjects with highest significance levels ($p < 0.01$).

Up-regulated miRNAs			Down-regulated miRNAs		
ID	**Fold Change**	***p* value**	**ID**	**Fold Change**	***p* value**
hsa-miR-1587	25.79288964	6.1e-11	hsa-miR-3125	0.002133045	1.8e-10
hsa-miR-223-3p	22.32543862	6.4e-05	hsa-miR-642b-3p	0.003146299	7.3e-13
hsa-miR-1915-3p	20.29551145	7.0e-13	hsa-miR-4505	0.005873541	1.0e-10
hsa-miR-6125	19.50742743	6.1e-11	hsa-miR-498	0.007903146	1.0e-11
hsa-miR-1273g-3p	12.49108029	2.0e-10	hsa-miR-874	0.009242724	3.7e-12
hsa-miR-3135b	10.76713107	2.2e-06	hsa-miR-4714-3p	0.011893235	1.6e-09
hsa-miR-6068	9.931160484	1.7e-08	hsa-miR-3177-5p	0.013058702	3.7e-12
hsa-miR-5001-5p	9.693932693	1.0e-08	hsa-miR-486-5p	0.035139858	3.5e-10
hsa-miR-4787-5p	9.57855506	1.7e-08	hsa-miR-4314	0.04792918	9.3e-11
hsa-miR-4687-3p	9.208735521	7.0e-10	hsa-miR-4516	0.050084115	7.3e-13
hsa-miR-4459	7.51622894	2.3e-03	hsa-miR-3141	0.056298912	4.8e-06
hsa-miR-6089	7.130171895	1.4e-04	hsa-miR-631	0.079323159	3.2e-10
hsa-miR-371b-5p	7.064802086	7.2e-07	hsa-miR-345-3p	0.0815455	3.4e-07
hsa-miR-4507	6.798294905	8.1e-08	hsa-miR-6088	0.090844785	5.0e-09
hsa-miR-23a-3p	6.768617929	5.6e-05	hsa-miR-4520b-3p	0.103252363	3.2e-10
hsa-miR-6724-5p	6.56447502	1.3e-08	hsa-miR-642a-3p	0.110994127	2.7e-08

hsa-miR-4466	6.290714855	2.4e-07	hsa-miR-4304	0.113318287	3.3e-11
hsa-miR-638	5.81701254	2.4e-07	hsa-miR-1233-1-5p	0.115027055	2.8e-11
hsa-miR-15b-5p	5.093308711	2.9e-03	hsa-miR-4802-3p	0.120586864	6.1e-09
hsa-let-7f-5p	4.710857854	1.9e-03	hsa-miR-302d-5p	0.137088177	4.3e-09

MiRNAs target identification

The number of target genes which were identified for up-regulated and down-regulated DE-miRNAs in patients with prostate cancer vs. healthy individuals were 1326 and 1901, respectively. To verify that these genes are functionally interrelated, gene-gene interaction networks were extracted for PCa-associated genes (Figure 2). These networks indicated hub genes among the genes encode for positively or negatively regulated vesicle-free microRNAs in PCa. Full sets of dysregulated target genes were screened to detect functional profile of PCa DE-miRNAs. MiRNA-gene interaction networks were separately obtained for up- and down-regulated DE-miRNAs (Figure 3). Network analysis revealed the most contributing microRNAs and genes which are responsible for development of prostate cancer. Table 2 presents these network items with their relevant topological parameters.

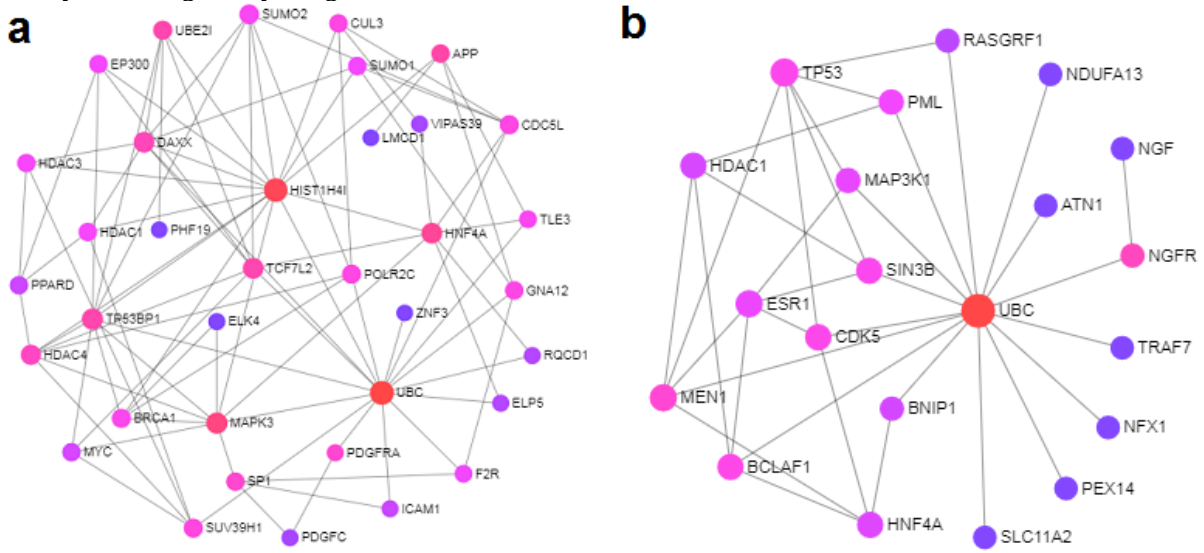

Figure 2. Interaction networks constructed for (a) positively, and (b) negatively regulated miRNA target genes in PCa samples. Data was obtained from MirWalk online tool.

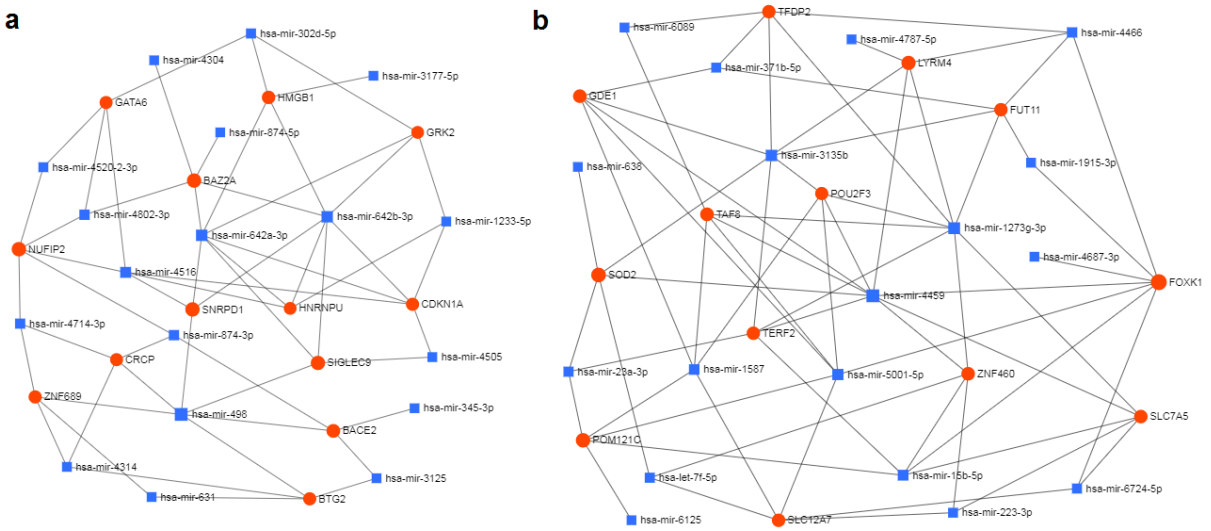

Figure 3. MiRNA-gene interaction networks which were built for (a) up-regulated, and (b) down-regulated differentially expressed microRNAs in prostate cancer samples are shown. miRNet 2.0 was utilized for extracting and analyzing the miRNA-gene networks.

Table 2. MicroRNAs which are involved in the regulatory network of PCa development in addition to their target genes encode for these miRs are shown in the table.

Node item	Degree	Betweenness	Node item	Degree	Betweenness
Up-regulated miRs			**Down-regulated miRs**		
hsa-let-7f-5p	3	8.62873	hsa-miR-498	6	143.32320
hsa-miR-23a-3p	3	10.17321	hsa-miR-631	2	1.15202
hsa-miR-223-3p	3	3.53887	hsa-miR-302d-5p	3	12.16220
hsa-miR-15b-5p	5	27.67949	hsa-miR-3125	2	2.68535
hsa-miR-638	1	0	hsa-miR-4304	1	0
hsa-miR-1915-3p	2	2.88528	hsa-miR-4314	3	4.13384
hsa-miR-4459	9	82.29159	hsa-miR-642b-3p	7	76.55934
hsa-miR-3135b	7	44.17389	hsa-miR-4505	2	2.69919
hsa-miR-4466	4	15.77690	hsa-miR-4516	5	54.41083
hsa-miR-1587	5	21.67443	hsa-miR-3177-5p	1	0
hsa-miR-4687-3p	1	0	hsa-miR-4714-3p	3	20.77789
hsa-miR-371b-5p	3	4.10794	hsa-miR-4802-3p	3	28.51971
hsa-miR-4787-5p	1	0	hsa-miR-642a-3p	7	76.55934
hsa-miR-5001-5p	6	39.84926	hsa-miR-345-3p	1	0
hsa-miR-1273g-3p	8	55.44037	hsa-miR-874-3p	3	25.92176
hsa-miR-6089	2	1.58383	hsa-miR-4520-2-3p	2	4.81212
hsa-miR-6125	1	0	hsa-miR-1233-5p	3	2.28321
hsa-miR-6724-5p	3	7.19620	hsa-miR-874-5p	1	0
Target genes			**Target genes**		
SOD2	5	46.16845	SNRPD1	4	70.67985
SLC12A7	5	22.01802	BTG2	4	28.80109
ZNF460	5	25.13426	BACE2	4	48.91486
TERF2	5	21.99524	SIGLEC9	4	66.69920
POM121C	5	40.97833	CRCP	4	22.87670
SLC7A5	5	23.57412	ZNF689	4	29.71133
FOXK1	7	71.76947	GRK2	4	15.79189
FUT11	5	21.23700	GATA6	4	25.11072
POU2F3	5	16.91934	HMGB1	4	37.30794
GDE1	5	25.12142	BAZ2A	5	68.81256
LYRM4	5	36.66783	CDKN1A	5	28.53590
TAF8	5	27.21987	HNRNPU	4	15.30749
TFDP2	5	21.19667	NUFIP2	5	72.45046

Functional enrichment of target genes

We carried out functional annotation analysis based on the GO of BP components to investigate the roles of microRNAs target genes. Results confirmed that miRs with significantly changed expression levels between disease and healthy conditions could alter some key biological processes and thus, promote manifestation of the disease.

Among various biological processes, target genes of up-regulated miRNAs were mostly enriched in the following pathways ($p < 0.01$): enzyme linked receptor protein signaling pathway, regulation of apoptosis, negative regulation of cytoskeleton organization, transmembrane receptor protein tyrosine kinase signaling pathway, regulation of cellular component size, negative regulation of protein complex disassembly, and transforming growth factor beta receptor signaling pathway. The most important prostate cancer related GO biological processes were considered as the ones with the most number of involved genes (Figure 4a). On the other hand, target genes of down-regulated

miRNAs were shown to take part in negative regulation of megakaryocyte differentiation, telomere capping and organization, beta-catenin-TCF complex assembly, nucleosome assembly, chromatin silencing at rDNA, epigenetic regulation of gene expression, transcription initiation, double-strand break repair via nonhomologous end joining, protein hetero tetramerization, cellular protein metabolic process, gene silencing by RNA, semaphorin-plexin signaling pathway, histone H3 deacetylation, positive regulation of phosphatidylinositol 3-kinase signaling, and regulation of cell shape ($p < 0.01$ for all). The relative contributions of significant GO terms in negative regulation of PCa-associated miRs is shown in Figure 4b.

The most frequently represented genes which were determined based on the total count of the genes in the processes are: Promyelocytic leukemia (PML), Forkhead box K1 (FOXK1), Superoxide dismutase 2 (SOD2), Solute carrier family 12 member 7 (SLC12A7), POU class 2 homeobox 3 (POU2F3), Zinc finger protein 460 (ZNF460), Telomeric repeat binding factor 2 (TERF2), NADH: ubiquinone oxidoreductase subunit A13 (NDUFA13), Cyclin dependent kinase 5 (CDK5), Mitogen-activated protein kinase kinase kinase 1 (MAP3K1), and Maelstrom spermatogenic transposon silencer (MAEL). Both highly represented genes and annotated genes of the miRNA-gene interaction network were used to draw a GO term-gene map for up-regulated microRNAs (Figure 5a).

In case of down-regulated genes, total counts of target genes in the processes revealed Histone H4 family members, High mobility group box 1 (HMGB1), Cell division cycle 5 like (CDC5L), Death domain associated protein (DAXX), Intercellular adhesion molecule 1 (ICAM1), Small nuclear ribonucleoprotein D1 polypeptide (SNRPD1), Zinc finger protein 689 (ZNF689), Heterogeneous nuclear ribonucleoprotein U (HNRNPU), Cyclin dependent kinase inhibitor 1A (CDKN1A), BTG anti-proliferation factor 2 (BTG2), Bromodomain adjacent to zinc finger domain 2A (BAZ2A), Platelet derived growth factor receptor alpha (PDGFRA), Peroxisome proliferator activated receptor delta (PPARD), Mitogen-activated protein kinase 3 (MAPK3), Tumor protein p53 binding protein 1 (TP53BP1), and zinc finger protein 3 (ZNF3) as the mostly represented genes. Both highly represented genes and annotated genes of the miRNA-gene interaction network were used to draw a GO term-gene map for down-regulated microRNAs (Figure 5b).

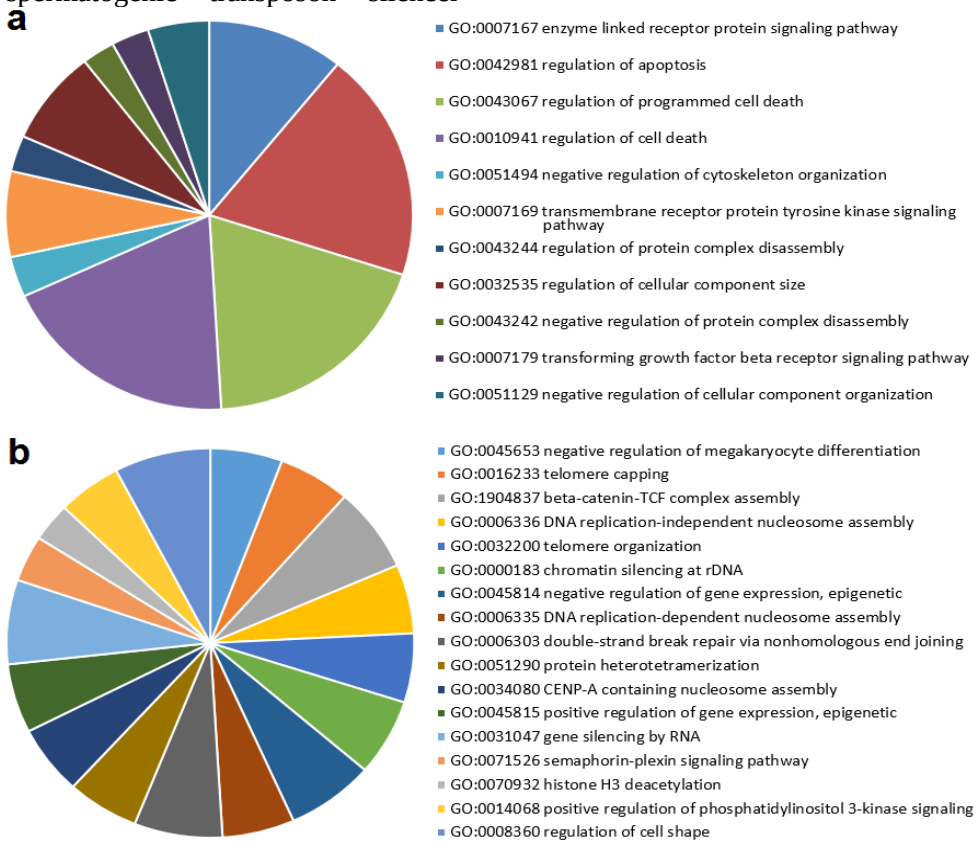

Figure 4. Pie charts indicated an overall view from biological processes which are associated with predicted miRs target genes, for (a) up-regulated, and (b) down-regulated microRNAs, respectively. Each sector size is proportional to the number of genes which are involved in the GO biological process ($p < 0.01$).

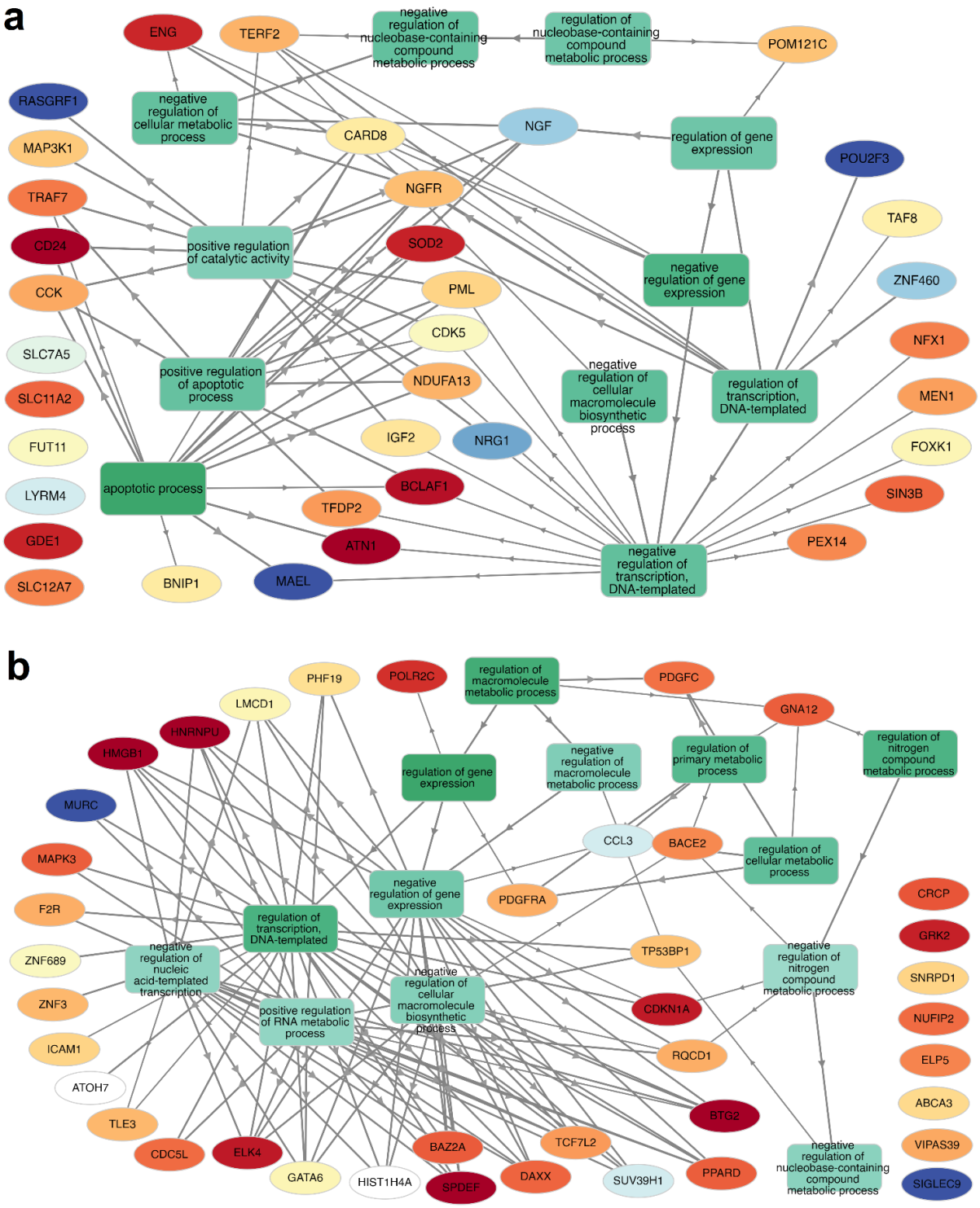

Figure 5. Networks of significant gene targets of (a) up-regulated, and (b) down-regulated differentially expressed miRNAs. The intensity of GO term node colors indicates the significance of enrichment ($p < 0.0001$). Genes were colored due to their expression levels in the prostate gland tissue; blue represents the lowest level of gene expression, while, genes with the highest level of expression are shown with red color. To visualize the functional interconnection of target genes and their significant GO terms, the GOnet server tool was applied. In the server, *p* value threshold for relation significance was set 1.94e-5.

Discussion

To date, numerous microRNAs have been identified in human prostate tissue, serum, plasma, and urine which are associated with localized and metastatic prostate cancers. Nevertheless, further experiments are required to evaluate the potential of miRNAs as biomarkers for the early diagnosis of PCa, particularly among men with increased risk for prostate cancer (McDonald et al., 2018). In the present study, we investigated whether blood circulating microRNAs demonstrate differential expressions according to the status of prostate cancer. As we were endeavored to find proper miRs with diagnostic value, vesicle-free microRNAs were selected due to their beneficial applicable properties in comparison to the vesicle-enclosed miRNAs. This kind of studies not only will help us to distinguish individuals with urgent need for performing prostate biopsy, but also may elucidate factors which are involved in the pathogenesis of the prostate cancer. Accordingly, the up-regulation of miR-375 and down-regulation of miR-146a-5p as extracellular vesicle-incorporated miRs with highly established roles in the pathogenesis of prostate cancer, were not the matter of our analyses (Endzelins et al., 2017). This approach will help us to avoid increased variability and more conveniently translate our findings to the clinical practices (Abramovic et al., 2020; Fabris et al., 2016; Fendler et al., 2016; Fredsoe et al., 2020).

Although we applied GEO series from the previous study performed by Mello-Grand et al. (2019), we only included PCa and healthy samples in the comparisons and omitted benign prostatic hyperplasia (BPH) samples from the analysis. In addition, we were stricter in the selection of the filters, as we considered p values <0.01 as the statistically significant differences. Moreover, despite the original research we did not assign any cut-off value during our analysis due to the below mentioned reasons: (1) we aimed to find miRs biomarkers that could discriminate PCa cases from healthy controls; (2) we aimed to find the responsible genes and their key interactions which are involved in the PCa pathogenesis using the corresponding GEO series accession. As an additional comparison, Mello-Grand et al. (2019) did not consider the target genes profiles for the miRs identified in their research. Also, they did not perform network and functional enrichment analyses.

The relationship between identified miRs and their target genes with prostate cancer has sufficient evidences from the literature. Significant up-regulation of the central miRNAs such as miR-223-3p (Bahtiyar et al., 2018; Cao et al., 2015; Feng et al., 2018; Jiang et al., 2005; Kurozumi et al., 2016; Liu et al., 2018; Mi et al., 2007; Triozzi et al., 2012; Volinia et al., 2006; Wei et al., 2014), miR-23a-3p (Aghaee-Bakhtiari et al., 2015; Cai et al., 2015; Jiang et al., 2005; Mi et al., 2007; Porkka et al., 2007; Wen et al., 2015), miR-15b-5p (Chen et al., 2018; Musumeci et al., 2011), miR-let-7f-5p (Ge et al., 2020; Jiang et al., 2005; Mello-Grand et al., 2019; Porkka et al., 2007), and miR-371 (HMDD v.3.2) in PCa tumors has been confirmed previously. Furthermore, miR-642b (HMDD v.3.2), miR-498 (Bendoraite et al., 2010; Porkka et al., 2007), miR-486 (Navon et al., 2009; Song et al., 2015), miR-4516 (HMDD v.3.2), miR-631 (HMDD v.3.2), miR-345 (Chen et al., 2016; Jiang et al., 2005; Porkka et al., 2007; Sayed et al., 2007; Tinay et al., 2018; Wang et al., 2014), miR-642a (HMDD v.3.2), and miR-302d (HMDD v.3.2) have been shown to be significantly down-regulated in PCa samples. These cell-free miRNAs could be considered as biomarkers for early detection of PCa (Zhang et al., 2014).

Combined analysis of the expression patterns and network topologies for identified DE-miRNAs in the present study allowed the identification of the miRs with high diagnostic potentials. Interestingly such potential markers can be found only among positively regulated miRs, including miR-4459, miR-1273g-3p, miR-3135b, miR-5001-5p, and miR-1587.

The miRNA-gene interaction networks constructed for DE-miRNAs confirmed most of the target genes which were predicted for these DE-microRNAs. Functional annotation of target genes indicated that the up-regulation of miRs in PCa is responsible for suppressed apoptotic process, enhanced nucleobase-containing metabolic process, and enhanced gene expressions which imply to augmented cell proliferation. The most important genes were FOXK1 (Chen et al., 2017), PML (Gurrieri et al., 2004), CD24 (Rouhrazi et al., 2018), ATN1 (Maugham et al., 2017), BCALF1, ENG (Liu et al., 2002), and SOD2 (Burri et al., 2008). In contrast, down-regulation of the identified miRs resulted in the negative regulation of gene expressions and macromolecule biosynthesis, implying that they are tumor suppressive microRNAs. The most important genes were BAZ2A (Gu et al., 2015), CDKN1A (Mitchell and El-Deiry, 1999), NUFIP2, HARNPU, HMGB1 (Wang et al., 2018), SPDEF (Findlay et al., 2008), BTG2 (Lim et al., 2008), and HIST1H4A (Yao et al., 2014).

Overall, we propose miR-4459, miR-1273g-3p, miR-3135b, miR-5001-5p, and miR-1587 as novel

microRNAs with great potential to be utilized as diagnostic biomarkers for distinguishing prostate cancer patients from healthy individuals. Although, these *in silico* findings should be verified through experimental studies.

Conflict of interest

The authors declare that there is no conflict of interest.

References

Abramovic I., Ulamec M., Bojanac A. K., Bulic-Jakus F., Jezek D. and Sincic N. (2020) miRNA in prostate cancer: challenges toward translation. Epigenomics 12:543-558.

Aghaee-Bakhtiari S. H., Arefian E., Naderi M., Noorbakhsh F., Nodouzi V., Asgari M., Fard-Esfahani P., Mahdian R. and Soleimani M. (2015) MAPK and JAK/STAT pathways targeted by miR-23a and miR-23b in prostate cancer: computational and in vitro approaches. Tumour Biology 36:4203-4212.

Bahtiyar N., Onaran I., Aydemir B., Baykara O., Toplan S., Agaoglu F. Y. and Akyolcu M. C. (2018) Monitoring of platelet function parameters and microRNA expression levels in patients with prostate cancer treated with volumetric modulated arc radiotherapy. Oncology Letters 16:4745-4753.

Bendoraite A., Knouf E. C., Garg K. S., Parkin R. K., Kroh E. M., O'Briant K. C., Ventura A. P., Godwin A. K., Karlan B. Y., Drescher C. W., Urban N., Knudsen B. S. and Tewari M. (2010) Regulation of miR-200 family microRNAs and ZEB transcription factors in ovarian cancer: evidence supporting a mesothelial-to-epithelial transition. Gynecologic Oncology 116:117-125.

Boesen L., Norgaard N., Logager V., Balslev I., Bisbjerg R., Thestrup K. C., Winther M. D., Jakobsen H. and Thomsen H. S. (2018) Assessment of the diagnostic accuracy of biparametric magnetic resonance imaging for prostate cancer in biopsy-naive men: the biparametric MRI for detection of prostate cancer (BIDOC) study. JAMA Network Open 1: e180219.

Bolla M. and van Poppel H. (2012) Management of prostate cancer: a multidisciplinary approach. Springer-Verlag, Berlin Heidelberg. VIII, 334.

Bray F., Ferlay J., Soerjomataram I., Siegel R. L., Torre L. A. and Jemal A. (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal of Clinicians 68:394-424.

Burri R. J., Stock R. G., Cesaretti J. A., Atencio D. P., Peters S., Peters C. A., Fan G., Stone N. N., Ostrer H. and Rosenstein B. S. (2008) Association of single nucleotide polymorphisms in SOD2, XRCC1 and XRCC3 with susceptibility for the development of adverse effects resulting from radiotherapy for prostate cancer. Radiation Research 170:49-59.

Cai S., Chen R., Li X., Cai Y., Ye Z., Li S., Li J., Huang H., Peng S., Wang J., Tao Y., Wen X., Mo J., Deng Z., Zhang Y. and Gao X. (2015) Downregulation of microRNA-23a suppresses prostate cancer metastasis by targeting the PAK6-LIMK1 signaling pathway. Oncotarget 6:3904-3917.

Cao L., Zhang X., Cao F., Wang Y., Shen Y., Yang C., Uzan G., Peng B. and Zhang D. (2015) Inhibiting inducible miR-223 further reduces viable cells in human cancer cell lines MCF-7 and PC3 treated by celastrol. BMC Cancer 15:873.

Chang L., Zhou G., Soufan O. and Xia J. (2020) miRNet 2.0 - network-based visual analytics for miRNA functional analysis and systems biology. Nucleic Acids Research 48:W244-W251.

Chen D., Wang K., Li X., Jiang M., Ni L., Xu B., Chu Y., Wang W., Wang H., Kang H., Wu K., Liang J. and Ren G. (2017) FOXK1 plays an oncogenic role in the development of esophageal cancer. Biochemical and Biophysical Research Communications 494:88-94.

Chen Q. G., Zhou W., Han T., Du S. Q., Li Z. H., Zhang Z., Shan G. Y. and Kong C. Z. (2016) MiR-345 suppresses proliferation, migration and invasion by targeting Smad1 in human prostate cancer. Journal of Cancer Research and Clinical Oncology 142:213-224.

Chen R., Sheng L., Zhang H. J., Ji M. and Qian W. Q. (2018) miR-15b-5p facilitates the tumorigenicity by targeting RECK and predicts tumour recurrence in prostate cancer. Journal of Cellular and Molecular Medicine 22:1855-1863.

Elkjaer M. C., Andersen M. H., Hoyer S., Pedersen B. G. and Borre M. (2018) Multi-parametric magnetic resonance imaging monitoring patients in active surveillance for prostate cancer: a prospective cohort study. Scandinavian Journal of Urology 52:8-13.

Endzelins E., Berger A., Melne V., Bajo-Santos C., Sobolevska K., Abols A., Rodriguez M., Santare D., Rudnickiha A., Lietuvietis V., Llorente A. and Line

A. (2017) Detection of circulating miRNAs: comparative analysis of extracellular vesicle-incorporated miRNAs and cell-free miRNAs in whole plasma of prostate cancer patients. BMC Cancer 17:730.

Fabris L., Ceder Y., Chinnaiyan A. M., Jenster G. W., Sorensen K. D., Tomlins S., Visakorpi T. and Calin G. A. (2016) The potential of microRNAs as prostate cancer biomarkers. European Urology 70:312-322.

Fendler A., Stephan C., Yousef G. M., Kristiansen G. and Jung K. (2016) The translational potential of microRNAs as biofluid markers of urological tumours. Nature Reviews in Urology 13:734-752.

Feng Q., He P. and Wang Y. (2018) MicroRNA-223-3p regulates cell chemo-sensitivity by targeting FOXO3 in prostatic cancer. Gene 658:152-158.

Ferracin M., Veronese A. and Negrini M. (2010) Micromarkers: miRNAs in cancer diagnosis and prognosis. Expert Reviews of Molecular Diagnostics 10:297-308.

Findlay V. J., Turner D. P., Moussa O. and Watson D. K. (2008) MicroRNA-mediated inhibition of prostate-derived Ets factor messenger RNA translation affects prostate-derived Ets factor regulatory networks in human breast cancer. Cancer Research 68:8499-8506.

Fredsoe J., Rasmussen A. K. I., Mouritzen P., Bjerre M. T., Ostergren P., Fode M., Borre M. and Sorensen K. D. (2020) Profiling of circulating microRNAs in prostate cancer reveals diagnostic biomarker potential. Diagnostics (Basel) 10:188.

Ge Y., Wang Q., Shao W., Zhao Y., Shi Q., Yuan Q. and Cui L. (2020) Circulating let-7f-5p improve risk prediction of prostate cancer in patients with benign prostatic hyperplasia. Journal of Cancer 11:4542-4549.

Glinge C., Clauss S., Boddum K., Jabbari R., Jabbari J., Risgaard B., Tomsits P., Hildebrand B., Kaab S., Wakili R., Jespersen T. and Tfelt-Hansen J. (2017) Stability of circulating blood-based microRNAs - pre-analytic methodological considerations. PLoS One 12: e0167969.

Gu L., Frommel S. C., Oakes C. C., Simon R., Grupp K., Gerig C. Y., Bar D., Robinson M. D., Baer C., Weiss M., Gu Z., Schapira M., Kuner R., Sultmann H., Provenzano M., Yaspo M. L., Brors B., Korbel J., Schlomm T., Sauter G., Eils R., Plass C. and Santoro R. (2015) BAZ2A (TIP5) is involved in epigenetic alterations in prostate cancer and its overexpression predicts disease recurrence. Nature Genetics 47:22-30.

Gurrieri C., Capodieci P., Bernardi R., Scaglioni P. P., Nafa K., Rush L. J., Verbel D. A., Cordon-Cardo C. and Pandolfi P. P. (2004) Loss of the tumor suppressor PML in human cancers of multiple histologic origins. Journal of the National Cancer Institute 96:269-279.

Hamdy F. C., Donovan J. L., Lane J. A., Mason M., Metcalfe C., Holding P., Davis M., Peters T. J., Turner E. L., Martin R. M., Oxley J., Robinson M., Staffurth J., Walsh E., Bollina P., Catto J., Doble A., Doherty A., Gillatt D., Kockelbergh R., Kynaston H., Paul A., Powell P., Prescott S., Rosario D. J., Rowe E. and Neal D. E. (2016) 10-Year outcomes after monitoring, surgery, or radiotherapy for localized prostate cancer. The New England Journal of Medicine 375:1415-1424.

Helgstrand J. T., Roder M. A., Klemann N., Toft B. G., Lichtensztajn D. Y., Brooks J. D., Brasso K., Vainer B. and Iversen P. (2018) Trends in incidence and 5-year mortality in men with newly diagnosed, metastatic prostate cancer-A population-based analysis of 2 national cohorts. Cancer 124:2931-2938.

Huang X., Yuan T., Tschannen M., Sun Z., Jacob H., Du M., Liang M., Dittmar R. L., Liu Y., Kohli M., Thibodeau S. N., Boardman L. and Wang L. (2013) Characterization of human plasma-derived exosomal RNAs by deep sequencing. BMC Genomics 14:319.

Huang Z., Shi J., Gao Y., Cui C., Zhang S., Li J., Zhou Y. and Cui Q. (2019) HMDD v3.0: a database for experimentally supported human microRNA-disease associations. Nucleic Acids Research 47: D1013-D1017.

Jiang J., Lee E. J., Gusev Y. and Schmittgen T. D. (2005) Real-time expression profiling of microRNA precursors in human cancer cell lines. Nucleic Acids Research 33:5394-5403.

Kurozumi A., Goto Y., Matsushita R., Fukumoto I., Kato M., Nishikawa R., Sakamoto S., Enokida H., Nakagawa M., Ichikawa T. and Seki N. (2016) Tumor-suppressive microRNA-223 inhibits cancer cell migration and invasion by targeting ITGA3/ITGB1 signaling in prostate cancer. Cancer Science 107:84-94.

Laursen E. B., Fredsoe J., Schmidt L., Strand S. H., Kristensen H., Rasmussen A. K. I., Daugaard T. F., Mouritzen P., Hoyer S., Kristensen G., Stroomberg H. V., Brasso K., Roder M. A., Borre M. and Sorensen K. D. (2019) Elevated miR-615-3p

expression predicts adverse clinical outcome and promotes proliferation and migration of prostate cancer cells. The American Journal of Pathology 189:2377-2388.

Lim Y. B., Park T. J. and Lim I. K. (2008) B cell translocation gene 2 enhances susceptibility of HeLa cells to doxorubicin-induced oxidative damage. Journal of Biological Chemistry 283:33110-33118.

Lin S. and Gregory R. I. (2015) MicroRNA biogenesis pathways in cancer. Nature Reviews in Cancer 15:321-333.

Liu R. S. C., Olkhov-Mitsel E., Jeyapala R., Zhao F., Commisso K., Klotz L., Loblaw A., Liu S. K., Vesprini D., Fleshner N. E. and Bapat B. (2018) Assessment of serum microRNA biomarkers to predict reclassification of prostate cancer in patients on active surveillance. Journal of Urology 199:1475-1481.

Liu Y., Jovanovic B., Pins M., Lee C. and Bergan R. C. (2002) Over expression of endoglin in human prostate cancer suppresses cell detachment, migration and invasion. Oncogene 21:8272-8281.

Maugham M. L., Thomas P. B., Crisp G. J., Philp L. K., Shah E. T., Herington A. C., Chen C., Gregory L. S., Nelson C. C., Seim I., Jeffery P. L. and Chopin L. K. (2017) Insights from engraftable immunodeficient mouse models of hyperinsulinaemia. Scientific Reports 7:491.

McDonald A. C., Vira M., Shen J., Sanda M., Raman J. D., Liao J., Patil D. and Taioli E. (2018) Circulating microRNAs in plasma as potential biomarkers for the early detection of prostate cancer. Prostate 78:411-418.

Mello-Grand M., Gregnanin I., Sacchetto L., Ostano P., Zitella A., Bottoni G., Oderda M., Marra G., Munegato S., Pardini B., Naccarati A., Gasparini M., Gontero P. and Chiorino G. (2019) Circulating microRNAs combined with PSA for accurate and non-invasive prostate cancer detection. Carcinogenesis 40:246-253.

Mi S., Lu J., Sun M., Li Z., Zhang H., Neilly M. B., Wang Y., Qian Z., Jin J., Zhang Y., Bohlander S. K., Le Beau M. M., Larson R. A., Golub T. R., Rowley J. D. and Chen J. (2007) MicroRNA expression signatures accurately discriminate acute lymphoblastic leukemia from acute myeloid leukemia. Proceedings of the National Academy of Sciences of the United States of America 104:19971-19976.

Mitchell K. O. and El-Deiry W. S. (1999) Overexpression of c-Myc inhibits p21WAF1/CIP1 expression and induces S-phase entry in 12-O-tetradecanoylphorbol-13-acetate (TPA)-sensitive human cancer cells. Cell Growth & Differentiation 10:223-230.

Movahedpour A., Ahmadi N., Ghasemi Y., Savardashtaki A. and Shabaninejad Z. (2019) Circulating microRNAs as potential diagnostic biomarkers and therapeutic targets in prostate cancer: current status and future perspectives. Journal of Cellular Biochemistry 120:16316-16329.

Musumeci M., Coppola V., Addario A., Patrizii M., Maugeri-Sacca M., Memeo L., Colarossi C., Francescangeli F., Biffoni M., Collura D., Giacobbe A., D'Urso L., Falchi M., Venneri M. A., Muto G., De Maria R. and Bonci D. (2011) Control of tumor and microenvironment cross-talk by miR-15a and miR-16 in prostate cancer. Oncogene 30:4231-4242.

Navon R., Wang H., Steinfeld I., Tsalenko A., Ben-Dor A. and Yakhini Z. (2009) Novel rank-based statistical methods reveal microRNAs with differential expression in multiple cancer types. PLoS One 4:e8003.

Pomaznoy M., Ha B. and Peters B. (2018) GOnet: a tool for interactive gene ontology analysis. BMC Bioinformatics 19:470.

Porkka K. P., Pfeiffer M. J., Waltering K. K., Vessella R. L., Tammela T. L. and Visakorpi T. (2007) MicroRNA expression profiling in prostate cancer. Cancer Research 67:6130-6135.

Rouhrazi H., Turgan N. and Oktem G. (2018) Zoledronic acid overcomes chemoresistance by sensitizing cancer stem cells to apoptosis. Biotechnic & Histochemistry 93:77-88.

Sayed D., Hong C., Chen I. Y., Lypowy J. and Abdellatif M. (2007) MicroRNAs play an essential role in the development of cardiac hypertrophy. Circulation Research 100:416-424.

Schmidt L., Fredsoe J., Kristensen H., Strand S. H., Rasmussen A., Hoyer S., Borre M., Mouritzen P., Orntoft T. and Sorensen K. D. (2018) Training and validation of a novel 4-miRNA ratio model (MiCaP) for prediction of postoperative outcome in prostate cancer patients. Annals of Oncology 29:2003-2009.

Song C., Chen H., Wang T., Zhang W., Ru G. and Lang J. (2015) Expression profile analysis of microRNAs in prostate cancer by next-generation sequencing. Prostate 75:500-516.

Tinay I., Tan M., Gui B., Werner L., Kibel A. S. and Jia L. (2018) Functional roles and potential clinical application of miRNA-345-5p in prostate cancer.

Prostate 78:927-937.

Triozzi P. L., Achberger S., Aldrich W., Elson P., Garcia J. and Dreicer R. (2012) Differential immunologic and microRNA effects of 2 dosing regimens of recombinant human granulocyte/macrophage colony stimulating factor. Journal of Immunotherapy 35:587-594.

Urabe F., Matsuzaki J., Yamamoto Y., Kimura T., Hara T., Ichikawa M., Takizawa S., Aoki Y., Niida S., Sakamoto H., Kato K., Egawa S., Fujimoto H. and Ochiya T. (2019) Large-scale circulating microRNA profiling for the liquid biopsy of prostate cancer. Clinical Cancer Research 25:3016-3025.

Valadi H., Ekstrom K., Bossios A., Sjostrand M., Lee J. J. and Lotvall J. O. (2007) Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nature Cell Biology 9:654-659.

Volinia S., Calin G. A., Liu C. G., Ambs S., Cimmino A., Petrocca F., Visone R., Iorio M., Roldo C., Ferracin M., Prueitt R. L., Yanaihara N., Lanza G., Scarpa A., Vecchione A., Negrini M., Harris C. C. and Croce C. M. (2006) A microRNA expression signature of human solid tumors defines cancer gene targets. Proceedings of the National Academy of Sciences of the United States of America 103:2257-2261.

Wang C., Peng G., Huang H., Liu F., Kong D. P., Dong K. Q., Dai L. H., Zhou Z., Wang K. J., Yang J., Cheng Y. Q., Gao X., Qu M., Wang H. R., Zhu F., Tian Q. Q., Liu D., Cao L., Cui X. G., Xu C. L., Xu D. F. and Sun Y. H. (2018) Blocking the feedback loop between neuroendocrine differentiation and macrophages improves the therapeutic effects of enzalutamide (MDV3100) on prostate cancer. Clinical Cancer Research 24:708-723.

Wang K., Zhang S., Weber J., Baxter D. and Galas D. J. (2010) Export of microRNAs and microRNA-protective protein by mammalian cells. Nucleic Acids Research 38:7248-7259.

Wang S. Y., Shiboski S., Belair C. D., Cooperberg M. R., Simko J. P., Stoppler H., Cowan J., Carroll P. R. and Blelloch R. (2014) miR-19, miR-345, miR-519c-5p serum levels predict adverse pathology in prostate cancer patients eligible for active surveillance. PLoS One 9: e98597.

Wei Y., Yang J., Yi L., Wang Y., Dong Z., Liu Z., Ou-yang S., Wu H., Zhong Z., Yin Z., Zhou K., Gao Y., Yan B. and Wang Z. (2014) MiR-223-3p targeting SEPT6 promotes the biological behavior of prostate cancer. Scientific Reports 4:7546.

Wen Y. C., Lee W. J., Tan P., Yang S. F., Hsiao M., Lee L. M. and Chien M. H. (2015) By inhibiting snail signaling and miR-23a-3p, osthole suppresses the EMT-mediated metastatic ability in prostate cancer. Oncotarget 6:21120-21136.

Yao L., Li Y., Du F., Han X., Li X., Niu Y., Ren S. and Sun Y. (2014) Histone H4 Lys 20 methyltransferase SET8 promotes androgen receptor-mediated transcription activation in prostate cancer. Biochemical and Biophysical Research Communications 450:692-696.

Zhang W., Zang J., Jing X., Sun Z., Yan W., Yang D., Shen B. and Guo F. (2014) Identification of candidate miRNA biomarkers from miRNA regulatory network with application to prostate cancer. Journal of Translational Medicine 12:66.

Zhou G., Soufan O., Ewald J., Hancock R. E. W., Basu N. and Xia J. (2019) NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. Nucleic Acids Research 47: W234-W241.

Journal of Cell and Molecular Research (2020) 12 (1), 22-32 DOI: 10.22067/jcmr.v12i1.87629

Research Article

Identification of Breast Cancer Associated Putative Functional Single Nucleotide Polymorphisms in the Iranian Population through *In Silico* Analyses

Vida Nadafi Sichani[1], Seyed Alireza Emami[2], Morteza Bitaraf Sani[3]*, Mohammadreza Nassiri[4], Vinod Gopalan[5]

[1] *Department of Biology, Faculty of Science, University of Science and Art, Yazd, Iran*
[2] *Department of Biology, Faculty of Basic Sciences, Shahrekord Branch, Islamic Azad University, Shahrekord, Iran*
[3] *Animal Science Research Department, Yazd Agricultural and Natural Resources Research and Education Center, Agricultural Research, Education & Extension Organization (AREEO), Yazd, Iran*
[4] *Recombinant Proteins Research Group, Research Institute of Biotechnology and Genetic Group, Ferdowsi University of Mashhad, Mashhad, Iran*
[5] *School of Medicine and Medical Science, Menzies Health Institute Queensland, Griffith University, Gold Coast, 4222*

Received 8 July 2020 Accepted 22 August 2020

Abstract

Previous studies have found several distinct alleles at both levels of transcriptional activity and protein-DNA binding manners in breast cancer patients vs. healthy individuals through multi-step experimental approaches. This study presents a computational-based model to investigate the regulatory potential and functional properties of disease-related non-coding single nucleotide polymorphisms (SNPs) variants through several online *in silico* tools in the Iranian population. The association between the risk of breast cancer and its putative single nucleotide polymorphisms in the Iranian population was investigated through SNPedia database and genome-wide association studies (GWAS). Furthermore, a meta-analysis was performed by Comprehensive Meta-Analysis (CMA) software. Functional analyses were carried out through LDlink, HaploReg, and RegulomeDB. The impact of each single nucleotide polymorphism on gene expression profiles and transcription factor binding sites were predicted by the RegulomeDB. "5", "6", and "1d" scores were assigned to rs3746444, rs1062577, and rs1049174 by this scoring system, respectively. RegulomeDB scores of rs3746444-*MYH7B/MIR499A* and rs1062577-*ESR1* suggested that they are not putative functional single nucleotide polymorphisms; and may not associate with significant eQTL signals. The "1d" score for rs1049174-*RP11-277P12.20* confirmed an association with the expression of the target gene. Proxy variants rs6088678 and rs2617160 have been identified using LDlink in non-coding segments. They were in strong linkage disequilibrium (LD) with single nucleotide polymorphisms rs3746444 and rs1049174, respectively. Also, non-coding variants rs6088678-*TRPC4AP* and rs2617160- *RP11-277P12.20* with high-ranked scores showed the strongest related-expression. This work provides a rapid and direct *in silico*-based approach for the identification of functional genetic variants in the breast cancer. These analyses were conducted to evaluate the association of intended SNPs with the regulatory elements of histones, DNases, motif changes, and selected eQTL signals. It can be extended to some other complex single nucleotide polymorphism-related diseases.

***Keywords*:** Epigenetics, Functional single nucleotide polymorphisms, Genome-Wide Association Studies, Linkage Disequilibrium, RegulomeDB scoring system.

Introduction

Single Nucleotide Polymorphisms (SNPs) represent the most common markers of the genome diversity among individuals (Coetzee et al., 2012). The overwhelming majority of significantly associated genetic variants identified through GWAS were drop down outside of the coding area. Hence, it is difficult to understand how specific SNP increases disease susceptibility (Meng et al., 2018). Single nucleotide polymorphisms have a crucial role in the prediction of the risk of various complicated diseases including cancer (Seyedmir et al., 2017). Cis-regulatory regions (non-coding DNA regions) comprise distal elements such as promoters, enhancers, and insulators, which regulate transcriptional activities and complex spatial and temporal gene expressions following the binding of

*Corresponding author's e-mail address: *m.bitaraf@areeo.ac.ir*

transcription factors (Bauer-Mehren et al., 2009).

In addition, the majority of epigenetic changes may be reversible or preventable. So, the restoration of epigenetic changes could be applied as a proper strategy for cancer treatment or prevention (Coetzee et al., 2012). There are highly advanced web-based tools with the capacity for the annotation of a specific SNP to a target gene. Also, it is possible to measure the causal risk among numerous non-coding loci before performing time-consuming validation experiments. Such experiments will enable us to accurately predict the likelihood of particular cancer risk for individuals or communities (Coetzee et al., 2012). These types of studies are based on two hypotheses: I) alterations in the regulatory areas are major determinants of gene expression modifications. II) motifs in regulatory regions exhibit a location preference (e.g. at the center of H3K27ac, H3K4me1 or DNase peak (Meng et al., 2018). It has been observed that the chromatin status of enhancers is determined by highly specific histone modification patterns which are strongly linked to cell-type-specific gene expression programs on a global scale.(Heintzman et al., 2009) Along with H3K4me1, a general signal for enhancers, H3K27ac enrichment is also dedicated to the identification of active enhancers. Sequences with high H3K4me1 enrichment, and low H3K27ac are considered as ready-to-activate enhancers and are associated with low gene expression levels (Rhie et al., 2013). Hence, in the present study, in line with these kind of experiments, a comprehensive in silico study was conducted based on the application of computational-based methods including RegulomeDB, HaploReg, and LDlink. Encyclopedia of DNA Elements (ENCODE, from ChIP-seq experiments), and Roadmap Epigenomics (from methods such as ChromHMM) were utilized as data resources (Edwards et al., 2013). We selected breast cancer as the phenotype of choice among others during genome-wide association studies (GWAS). A list of three breast cancer risk-associated SNPs was obtained from the GWAS Catalog and SNPedia. Our purpose was to determine the functional value of rs3746444, rs1062577, and rs1049174 SNPs which were obtained through wet-lab experiments in the Iranian population.

Indeed, we performed pairwise comparisons with functional proxy variants suggested by the LDlink application. LDlink (analysistools.nci.nih.gov) is a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants (Machiela et al., 2015). Due to the importance of linkage disequilibrium (LD) structures in the indigenous populations, LDlink was utilized for finding two proxy variants to be compared with query variants in the Iranian population. We found coding proxy variants with high RegulomeDB scores which do not have any functional effect on regulatory regions. On the other side, a non-coding proxy variant with low RegulomDB score was selected. RegulomeDB variant classification scheme is fully described by Boyle et al. (Boyle et al., 2012).

HaploReg v4.1 is another web-based tool for exploring the annotations and producing mechanistic hypotheses of the impact of noncoding variants on the clinical phenotypes and normal variations (Fayez et al., 2018).

Materials and Methods

Selection of SNPs

In the present study, SNPs associated with breast cancer risk in the Iranian population were identified through the SNPedia (www.snpedia.com) and GWAS Catalog (www.ebi.ac.uk›gwas). These detected SNPs include rs3746444 (Kabirizadehet al., 2015), (Jiang et al., 2015), (Zou et al., 2012), (Mu et al., 2017), (Wang et al., 2012), (Wang et al., 2012), rs1062577(Dehghan et al., 2017), (Chen et al., 2016), and rs1049174 (Ghobadzadeh et al., 2013). Different parameters including odds ratios (OD), confidence interval (CI), number of samples, author's name, and host countries for these SNPs were extracted from relevant literature to conduct a comprehensive meta-analysis. The best and most effective SNPs were selected for downstream procedures.

In-silico studies

LDlink (www.ldlink.nci.nih.gov), HaploReg (www.pubs.broadinstitute.org/mammals/haploreg/haploreg.pbp) and RegulomeDB (http://www.regulomedb.org) web tools and databases were applied to determine the functional value of desired polymorphisms. Figure 1 is outline of our processing pipeline.

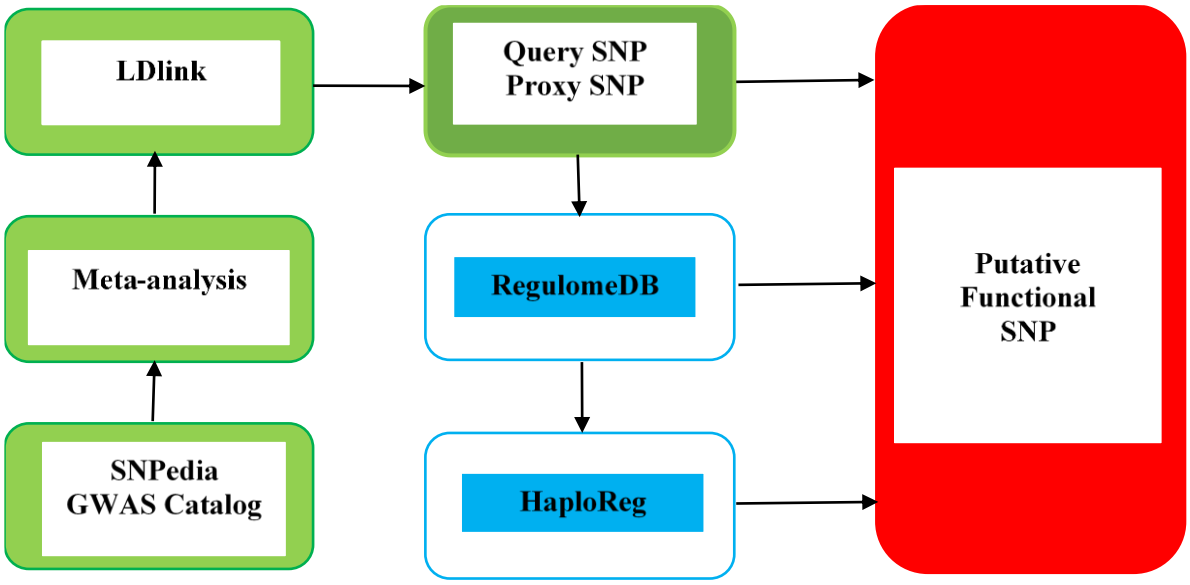

Figure 1. The pipeline consists of various key points including methods of SNP collection (SNPedia and GWAS Catalog), comprehensive meta-analysis (CMA), investigating the patterns of linkage disequilibrium across a variety of ancestral populations (LDlink), and developing the mechanistic hypothesis of the impact of non-coding variants on the clinical phenotypes (RegulomeDB and HaploReg). At the final step, we endeavored to confirm whether these SNPs are located in the regulatory segments and have functional impact on the gene expression patterns (Putative Functional SNP).
SNPedia: wiki-based bioinformatics web site that serves as a database of single nucleotide polymorphisms (SNPs). NHGRI-EBI GWAS Catalog: publicly available resource of Genome Wide Association Studies (GWAS) and their results. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. HaploReg: a tool for exploring annotations of the noncoding genome at variants on haplotype blocks, such as candidate regulatory SNPs at disease associated loci. RegulomeDB: a database that annotates SNPs with known and predicted regulatory elements in the intergenic regions of the H. sapiens genome. Known and predicted regulatory DNA elements include regions of DNase hypersensitivity, binding sites of transcription factors, and promoter regions that have been biochemically characterized to regulation transcription. Sources of these data include public datasets from GEO, the ENCODE project, and published literature. Query SNP: variant RS number - RS number for query variant. RS number must match a bi-allelic variant. Table of proxy variants: by default, the ten variants with the highest R2 values and closest distance to the query variant are displayed. External links lead to the variant RS number in dbSNP, coordinates in the UCSC Genome Browser, and regulatory information (if any) in RegulomeDB.

The LDlink web tool was used to detect proxy SNPs with strong LD (≥ 0.8) for rs3746444, rs1062577, and rs1049174. Proxy SNPs with these properties were selected for further analysis: I) coding SNPs with high RegulomeDB scores (4-6) and the least evidences for binding to regulatory proteins and participation in the gene expression regulation. II) non-coding SNPs with low RegulomeDB score (1a-1f) and the most evidences for binding to regulatory proteins and participation in the gene expression regulation. In addition, LD hap option of the LDlink was applied to evaluate haplotype frequencies between input SNPs and proxy SNPs.
Histone modifications in human tissues relevant to the breast cancer, such as breast myoepithelial primary cells (MEPs) and breast variant human mammary epithelial cells (vHMECs) (Fayez et al., 2018) were investigated by the HaploReg v4.1 tool. HaploReg was used to explore the annotations of the non-coding genome at variants on haplotype blocks, such as candidate regulatory SNPs at disease-related loci (Hamdi et al., 2018). The impact of genetic changes on different tissues and biological systems was revealed using HaploReg, RegulomeDB, and the genotype-tissue expression (GTEx) portal (Ward et al., 2011).

Results

Identification of single nucleotide polymorphisms

The meta-analysis results using CMA software are shown in the Table 1. The level of rs3746444 effect on the breast cancer predicted as "low". The interactive plot for rs3746444 represents the high density of SNPs with high LD around this variant (Figure 2).

Using LDlink web tool it was confirmed that two proxy SNPs (rs3746435 and rs6088678) with strong LD 0.87 with query SNP rs3746444, are associated with the breast cancer pathogenesis (Table 2). RegulomeDB score "5" was obtained for SNP rs3746435, located in the coding region. Meanwhile, based on this scoring system, low score of "1f" was measured for rs6088678 non-coding variant, implying a higher level of functional properties (Table 2).

It was confirmed by the application of HaploReg that rs3746444 induces histone modification H3K4me1_Enh in the breast myoepithelial primary cells and is located in the DNase I hypersensitive region of the breast variant of human mammary epithelial cells. It shows that motif changes may allow the DNase I to identify the available chromatin and cuts DNA at its respective region (Table 3). Although, proxy SNP rs3746435 (missense) with score "5", did not results in any histone modifications in the examined breast cancer cell lines (Table 3). It was also shown based on the results from HaploReg tool that proxy SNP rs6088678 caused the histone modification H3K9ac_ Pro. It is located within the promoter region and transcription start sites (TSS) and was effective in the regulation of *TRPC4AP* expression (Table 3). Query SNPs or SNPs previously identified as breast cancer associated ones in the Iranian population are marked in bold.

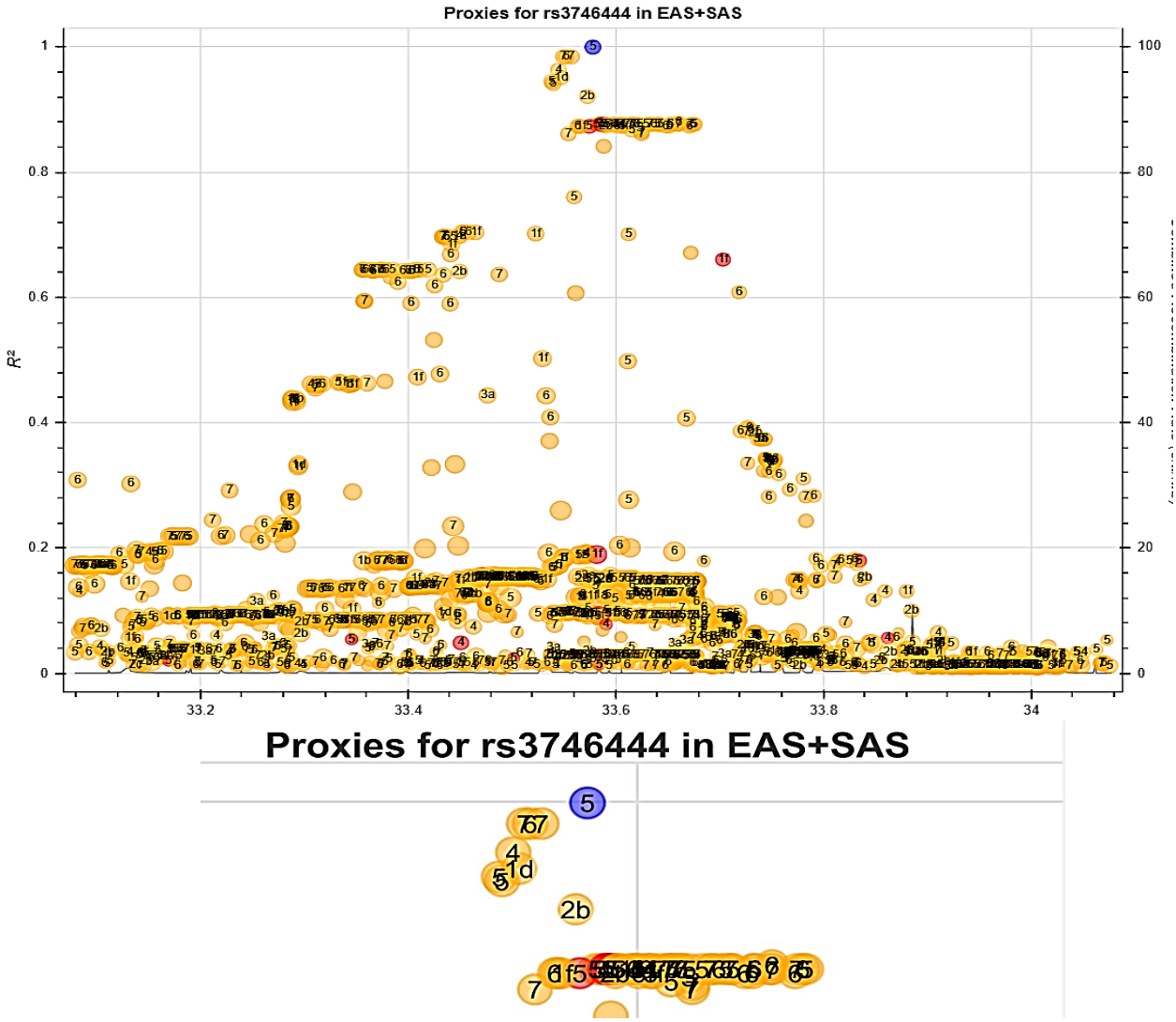

Figure 2. The interactive plot obtained for rs3746444(blue circle with RegulomeDB score=5) by the application of LDlink web tool. The complete and high-resolution chart could be viewed through the given link. (https://ldlink.nci.nih.gov/?var=rs3746444&pop=CHB%2BJPT%2BCHS%2BCDX%2BKHV%2BGIH%2BPJL%2BBEB%2BSTU%2BITU&r2_d=r2&window=500000&tab=ldproxy). Interactive plot: interactive plot of query variant(rs3746444) and all bi-allelic dbSNP variants plus or minus 500 kilobases (Kb) of the query variant(rs3746444). X axis is the chromosomal coordinates and the Y axis is the pairwise R2 value with the query variant as well as the

combined recombination rate from HapMap. Each point represents a proxy variant and is colored based on function, sized based on minor allele frequency, and labeled based on regulatory potential (regulatory potential number of rs3746444=5). Hovering over the point will display detailed information on the query and proxy variants. Reference population(s)((SAS) South Asian and (EAS) East Asian): selected from the drop-down menu. At least one 1000 Genomes Project sub-population is required, but more than one may be selected. R2/D' toggle: select if desired output is based on estimated R2 or D'.

Table 1. Results from meta-analysis of association studies for rs3746444.

Reference study	Odds ratios	Lower limit	Upper limit	z value	*p* value
Wang, Y., Yang, B. and Ren, X. (2012) Hsa-miR-499 polymorphism (rs3746444) and cancer risk: a meta-analysis of 17 case–control studies. Gene 509(2): 267-272.	1.230	1.059	1.429	2.710	0.007
Mu, K., Wu, Z. Z., Yu, J. P., Guo, W., Wu, N., Wei, L. J. and Liu, J. T. (2017) Meta-analysis of the association between three microRNA polymorphisms and breast cancer susceptibility. Oncotarget 8(40): 68809.	1.170	1.025	1.336	2.319	0.020
Wang, L., Qian, S., Zhi, H., Zhang, Y., Wang, B. and Lu, Z. (2012) The association between hsa-miR-499 T> C polymorphism and cancer risk: a meta-analysis. Gene 508(1): 9-14.	1.160	0.995	1.353	1.892	0.058
Zou, P., Zhao, L., Xu, H., Chen, P., Gu, A., Liu, N. and Lu, A. (2012) Hsa-mir-499 rs3746444 polymorphism and cancer risk: a meta-analysis. Journal of biomedical research 26(4): 253-259.	1.100	1.004	1.205	2.049	0.040
Jiang, S. G., Chen, L., Tang, J. H., Zhao, J. H. and Zhong, S. L. (2015) Lack of association between Hsa-Mir-499 rs3746444 polymorphism and cancer risk: meta-analysis findings. Asian Pacific Journal of Cancer Prevention 16(1):339-344.	1.180	1.035	1.346	2.466	0.014
Kabirizadeh, S., Azadeh, M., Mirhosseini, M., Ghaedi, K. and Tanha, H. M. (2016) The SNP rs3746444 within mir-499a is associated with breast cancer risk in Iranian population. Journal of Cellular Immunotherapy 2(2): 95-97.	1.922	1.064	3.471	2.167	0.030
	1.157	1.094	1.223	5.149	0.000

Data obtained by the LDlink revealed that rs3746444, which is located in a non-coding segment, indicated a RegulomeDB score "5" (Table 2). It seems that rs3746444 does not exhibit any significant biological activity such as alterations in the transcription factors binding capacity and gene regulatory effects in the Iranian population.

Table 2. Details of putative regulatory functions of query SNPs and their associated proxy SNPs.

cVariant	LD (r^2)	LD (D')	ASN freq	Enhancer histone marks	DNase	dbSNP func annot	GEN CODE genes	RegulomeDB score	Predicted function
rs3746444	1	1	0.17	IPSC, GI, MUS	BRST, BRN, LIV	Intronic	MYH7B & MIR499A	5	TF binding or DNase peak
rs3746435	0.87	1.0	0.12	SKIN, MUS	SKIN, PLCNT	Missense	MYH7B	5	TF binding or DNase peak
rs6088678	0.87	1.0	0.12	8 tissues	-	Intronic	TRPC4AP	1f	eQTL+TF binding / DNase peak
rs1062577	1	1	0.27	-	-	3'-UTR	ESR1	6	Motif hit
rs1049174	1	1	0.60	-	8 tissues	3'-UTR	RP11-277P12.20	1d	eQTL+TF binding+any motif+ DNase peak
rs2617160	0.88	0.9	0.59	10 tissues	5 tissues	Intronic	RP11-277P12.20	1f	eQTL+TF binding / DNase peak

A comparison of several factors between query and proxy variants with high LDs has been performed in the Asian population.

ASN freq: allele abundance in Asian population. SNP functional annotation: the functional area where mentioned SNP is located. GENCODE genes: the gene region in which SNP is located. RegulomeDB score: Regulome DB is a database

that scores SNPs functionality based upon experimental data. It is necessary to mention that in all tables Query SNPs are displayed in bold.

Table 3. Regulatory chromatin status from DNase and histone ChIP-Seq (Roadmap Epigenomics Consortium, 2015).

variant	Group	Description	H3K4me1	H3K4me3	H3K27ac	H3K9ac	DNase
rs3746444	Epithelial	Breast Myoepithelial Primary Cells	H3K4me1_Enh	-	-	-	-
rs3746444	Epithelial	Breast variant Human Mammary Epithelial Cells (vHMEC)	-	-	-	-	DNase
rs3746435	Epithelial	Breast Myoepithelial Primary Cells	-	-	-	-	-
rs3746435	Epithelial	vHMEC	-	-	-	-	-
rs6088678	Epithelial	Breast Myoepithelial Primary Cells	-	-	-	H3K9ac_Pro	-
rs1062577	Epithelial	vHMEC	H3K4me1_Enh	-	-	-	-
rs1049174	Epithelial	Breast Myoepithelial Primary Cells	H3K4me1_Enh	-	-	-	-
rs1049174	Epithelial	vHMEC	H3K4me1_Enh	-	-	-	-
rs2617160	Epithelial	Breast Myoepithelial Primary Cells	H3K4me1_Enh	-	-	-	-
rs2617160	Epithelial	vHMEC	H3K4me1_Enh	-	-	-	DNase

Query SNPs have been compared with proxy SNPs in terms of cellular and histological position. Histone modifications that each one creates in the target cells has been investigated.
Open chromatin: DNase1 hypersensitivity. Histone modifications: H3K4me1, H3K4me3, H3K9ac, H3K27ac. It is necessary to mention that in all tables Query SNPs are displayed in bold.

Increased frequency of haplotype AGC

The LD hap analysis (http://analysistools.nci.nih.gov/LDlink/tab=ldhap) showed increased frequency (79%) of AGC haplotype among three SNPs including rs3746444, rs3746435, and rs6088678. Results indicated that when the query SNP is adenine, the proxy allele for rs3746435 and rs6088678 will be G and C, respectively. There was a very strong LD among these three SNPs (87%). Also, the abundance of the AGC haplotype was high. It was revealed that allele A in query SNP rs3746444 is more likely to be associated with allele G; while if the query allele is G, it is likely that the proxy allele would be C (Figure 3).

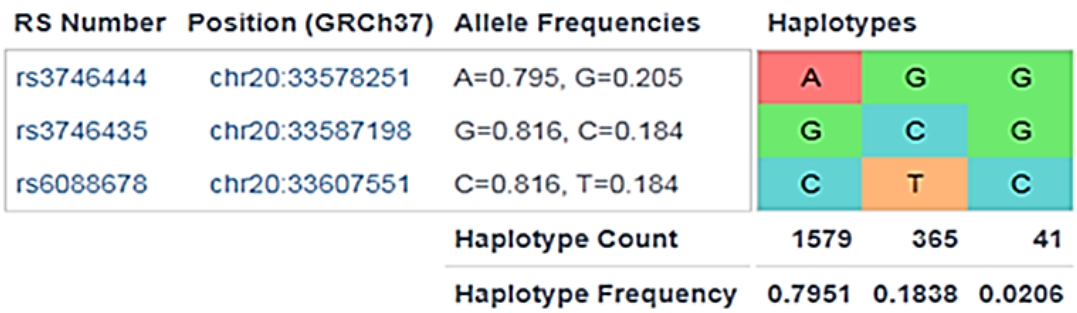

Figure 3. Haplotype Analysis of query SNP rs3746444 with two proxy SNP rs3746435 and rs6088678.
Results obtained from haplotype study of SNPs using LDlink web-based tool indicate that when the query SNP is adenine, the proxy allele for rs3746435 and rs6088678 will be G and C, respectively. There is a very strong LD among these three SNPs (87%).
(http://analysistools.nci.nih.gov/LDlink/tab=ldhap)

Association of SNPs with transcriptional levels of the target genes

Polymorphism rs3746444-*MYH7B/MIR499A* induces a poor transcriptional level in the breast MEPs and vHMECs, which in turn, will be resulted in the formation of a weak polycomb complex and reduced regulatory effects of the target gene. On the other side, rs3746435 induces a strong transcription in the examined cell lines. The proxy SNP rs6088678-*TRPC4AP* showed a strong transcription in addition to the score "1f" in both cell lines (Table 4). As demonstrated in Fig. 4, the non-coding proxy SNP rs6088678 with low score "1f", indicated the highest expression level in the breast tissue. Its value was equal to 40-60 Reads Per kilobase Million (RPKM) (Figure 4).

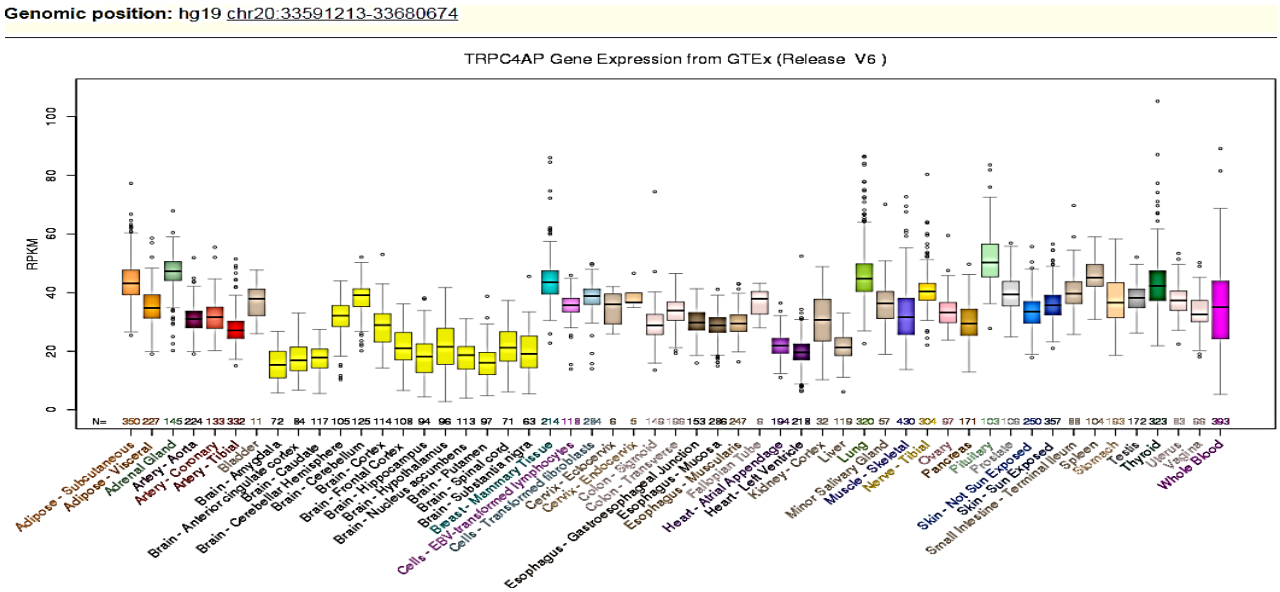

Figure 4. TRPC4AP gene expression from (GTEx) project for rs6088678.
The non-coding proxy SNP rs6088678 with low score "1f", indicates the highest expression level in the breast tissue (red arrow). Its value is equal to 40-60 Reads Per kilobase Million (RPKM).

Table 4. Genome browser, chromatin state and accessibility.

Method	SNP	Location	Chromatin State	Tissue Group	Tissue
ChromHMM	rs3746444	chr20:33575200..33578600	Weak transcription	Epithelial	Breast Myoepithelial Primary Cells
ChromHMM	rs3746444	chr20:33574000..33583000	Weak Repressed PolyComb	Epithelial	Breast variant Human Mammary Epithelial Cells (vHMEC)
ChromHMM	rs3746435	chr20:33583600..33645600	Strong transcription	Epithelial	Breast Myoepithelial Primary Cells
ChromHMM	rs3746435	chr20:33583000..33590400	Quiescent/Low	Epithelial	vHMEC
ChromHMM	rs6088678	chr20:33583600..33645600	Strong transcription	Epithelial	Breast Myoepithelial Primary Cells
ChromHMM	rs6088678	chr20:33603600..33608000	Strong transcription	Epithelial	vHMEC

ChromHMM	rs1062577	chr6:152398400..152431600	Quiescent/Low	Epithelial	Breast Myoepithelial Primary Cells
ChromHMM	rs1062577	chr6:152423600..152425200	Weak transcription	Epithelial	vHMEC
ChromHMM	rs1049174	chr12:10524400..10528200	Weak transcription	Epithelial	Breast Myoepithelial Primary Cells
ChromHMM	rs1049174	chr12:10525000..10525600	Enhancers	Epithelial	vHMEC
ChromHMM	rs2617160	chr12:10544600..10549000	Enhancers	Epithelial	Breast Myoepithelial Primary Cells
ChromHMM	rs2617160	chr12:10544800..10546400	Enhancers	Epithelial	vHMEC

ChromHMM (Hidden Markov Model) is applied to annotate the non-coding genome using epigenomic information between one or multiple cell types. Using RegulomeDB web-based tool, the transcription level of Query SNPs and proxy SNPs in different tissues and cell types has been determined. It is necessary to mention that in all tables Query SNPs are displayed in bold.

The NIH genotype-tissue expression (GTEx) project was created to establish sample and data resources for studies aimed to unravel the relationships between genetic variations and gene expression levels in multiple human tissues. This track shows median gene expression levels in 51 tissues and 2 cell lines, based on the RNA-seq data from the GTEx midpoint milestone data release (V6, October 2015). This release is formed based on the data from 8555 tissue samples obtained from 570 adult post-mortem individuals.

All the regulatory features which were seen in tables were obtained from ENCODE and NIH Roadmap Epigenomics data through the UCSC Genome Browser.

SNP rs1062577

Meta-analyses were not possible for query SNP rs1062577-*ESR1* due to the limited number of studies which carried out on rs1062577 in the Asia. The interactive plot in LDlink tool revealed that there is no SNP with strong LD (≥ 0.8) around rs1062577. RegulomeDB score of "6" revealed no remarkable effect on the gene expression levels (Table 2). However, it shows the target gene (*ESR1*) expression of rs1062577 in the breast cancer tissues. There are overwhelming data on the expression of *ESR1*, up to about 50 RPKM. Indeed, this polymorphism induces H3K4me1_Enh histone modification in vHMECs (Table 3). However, there was little evidences for rs1062577 to be a functional noncoding SNP.

SNP rs1049174 versus rs2617160

The interactive plot from LDlink tool, indicated low density of SNPs with strong LD around rs1049174. The non-coding proxy SNP rs2617160, located in the intronic region with score "1d", was selected for further analysis (Table 2). There was no coding SNP with strong LD for rs1049174. Both of the query SNP rs1049174, and proxy SNP rs2617160 caused H3K4me1_Enh histone modification in the investigated cell lines and were associated with the breast cancer tissue.

In contrast, proxy SNP rs2617160, thorough the induction of motif changes, produces open chromatin regions in vHMECs. Hence, DNase I can cut DNA in its respective region (Table 3). Both rs1049174, and 2617160 which were submerged in the RegulomeDB tool in addition to the proxy SNP rs2617160 are located in *RP11-277P12.20* enhancer sites of the examined breast cancer cell lines. rs1049174 caused a poor transcriptional level in the breast MEPs and is specifically located in the enhancer of the vHMECs (Table 4).

Discussion

Previous studies demonstrated that most of the GWAS variants fall in non-coding (nc) regions. The identification of the functions of these ncSNPs remains as a major challenge. The importance of understanding the functional contributions of specific risk variants to disease pathogenesis is widely accepted (Rhie et al., 2013). The biological effects of the most already studied SNPs in the Iranian population were not strong. In the present study through the application of a set of *in silico*

approaches, functional analyses were performed for previously known breast cancer risk associated SNPs in the Iranian population. The HaploReg database was established as a computer simulation tool by Ward and Kellis (Ward et al., 2011) to provide an intersects of single nucleotide variants (SNVs) with chromatin status (Ernst et al., 2010). For the first time, this work demonstrated that a comprehensive *in silico* analysis of well-known ncSNPs and regulatory regions is essential before we can attribute them to the Iranian population.

It was previously reported that rs3746444 (Kabirizadeh et al., 2016), rs1062577 (Dehghan et al., 2017), and rs1049174 (Ghobadzadeh et al., 2013) are associated with an increased risk of breast cancer in the Iranian population. We focused on non-coding proxy SNPs (LD≥0.8 with query SNPs rs3746444, rs1062577, and 1049174) which were obtained from LDlink. It was assumed that all non-coding variant SNPs which are located in the regulatory regions (promoter, enhancer, 5`UTR, 3`UTR) have a highly ranked RagulomeDB score (Table2). The meta-analysis of the rs3746444 in the Asian and Iranian population indicated a statistically significant relationship with the breast cancer by Odds Ratio(OR) = 1.15(1.09-1.22). These analyses were only possible for one SNP (Table 1).

Moreover, the regulatory effects of rs3746444-*MYH7B/MIR499A*, rs1062577-*ESR1*, and rs1049174-*RP11-277P12.20* and their related proxy SNPs were determined based on the high LD. We apply this analysis to identify the most likely functional variant among *MYH7B*, *ESR1*, and *RP11-277P12.20* genes. However, a solid framework of the functional significance of variants cannot be obtained by a single bioinformatics tool. Hence, some complementary tools were applied to perform the current study. Three computational-based tools including LDlink, HaploReg, and RegulomeDB were used for above mentioned SNPs in a combinatory mode to prioritized ncSNPs for their association with the disease status. The LD structure haplotype block for the Iranian population was not available because GWAS studies have not been performed previously in Iran. Hence, related information from the Asian population were utilized as a reference for LDlink studies.

We identified query SNP rs1049174 in 3'UTR region as the only previously wet-lab studied SNP with high ranked RegulomeDB score "1d" and validated functional effects (eQTL+TF binding+any motif+ DNase peak) (Table 2). rs1049174 caused histone modification H3K4me1 in both cancerous cell lines. It confirms that these enhancers are ready to be active.

The present study demonstrated that SNPs in the *MYH7B*, *TRPC4AP* and *RP11-277P12.20* genes (Table 2) in addition to the ncSNPs rs6088678, and rs2617160 are functionally important. Although, wet-lab experiments are essential for the validation of the results. Pairwise comparisons confirmed that intronic SNP rs6088678 (r^2 = 0.87 with rs3746444) and RegulomeDB score "1f" showed more evidences of being functional in comparison to rs3746444 (Table 2). It was shown that the rs6088678 induced histone modification H3K9ac in the breast myoepithelial primary cells (Table 3).

Due to our knowledge, this is the first association study between breast cancer susceptibility and polymorphisms of *MYH7B*, *MIR499A*, *TRPC4AP*, *ESR1* and *RP11-277P12.20* genes. These genes were selected using LDlink for the Iranian population. RegulomeDB is a powerful tool for the prediction of the regulatory potential of various variants. It is expected that the RegulomeDB web-based tool will be widely applied in the future for performing extensive association studies.

Conclusion

Considering the results of comparisons made in the present study which confirmed epigenetic properties for non-coding SNPs, the importance of these segments in the functional epigenetic studies were highlighted. Non-coding SNPs have a great impact on the binding capacity of regulatory proteins and gene expression pattern modifications as they can lead to histone modifications (Khurana et al., 2016). In order to evaluate the possible functional properties of shortlisted SNPs in the Iranian population, *in silico* analyses using LDlink, RegulomeDB and HaploReg are strongly recommended. It could be expected that our computational model could prioritize variants in the regulatory regions. Thus, it helps researchers to figure out functional variants of noncoding regions with key effects in the pathogenesis of various diseases.

Acknowledgments

The present article is part of the master's thesis of Biology and supported by the University of Science and Art. We also appreciate the efforts of Yazd Agricultural and Natural Resources Research Center members.

Conflict of interest

The authors declared no conflicts of interest.

References

Bauer-Mehren, A., Furlong, L. I., Rautschka, M. and Sanz, F. (2009) From SNPs to pathways: integration of functional effect of sequence variations on models of cell signalling pathways. BMC bioinformatics 10: S6.

Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M. and Cherry, J. M. (2012) Annotation of functional variation in personal genomes using RegulomeDB. Genome research 22(9): 1790-1797.

Chen, L., Kang, H., Jin, G. J., Chen, X., Zhang, Q. Y., Lao, W. T. and Li, R. (2016) The association between a novel polymorphism (rs1062577) in ESR1 and breast cancer susceptibility in the Han Chinese women. Gynecological Endocrinology 32(7): 553-556.

Coetzee, S. G., Rhie, S. K., Berman, B. P., Coetzee, G. A. and Noushmehr, H. (2012) FunciSNP: an R/bioconductor tool integrating functional non-coding data sets with genetic association studies to identify candidate regulatory SNPs. Nucleic acids research 40(18): e139-e139.

Dehghan, Z., Sadeghi, S., Tabatabaeian, H., Ghaedi, K., Azadeh, M., Fazilati, M. and Bagheri, F. (2017) ESR1 single nucleotide polymorphism rs1062577 (c.* 3804T> A) alters the susceptibility of breast cancer risk in Iranian population. Gene 611: 9-14.

Edwards, S. L., Beesley, J., French, J. D. and Dunning, A. M. (2013) Beyond GWASs: illuminating the dark road from association to function. The American Journal of Human Genetics 93(5): 779-797.

Ernst, J. and Kellis, M. (2010) Discovery and characterization of chromatin states for systematic annotation of the human genome. Nature biotechnology 28(8): 817-825.

Fayez, A. (2018) Using postgenome-wide association study analysis; Vars2-Pic3ca-AKT is novel putative interactive pathway associated with conotruncal heart defects. Biomedical and Biotechnology Research Journal 2(4): 269.

Ghobadzadeh, S., Shams, A., Eslami, G. and Mirghanizadeh, A. (2013) Investigation of NKG2D rs1049174G> C Gene Polymorphism in Women with Breast Cancer. SSU_Journals 21(3): 291-299.

Hamdi, Y., Rekaya, M. B., Jingxuan, S., Nagara, M., Messaoud, O., Elgaaied, A. B. and Boussen, H. (2018) A genome wide SNP genotyping study in the Tunisian population: specific reporting on a subset of common breast cancer risk loci. BMC cancer 18(1): 1-14.

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F. and Ching, K. A. (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459(7243): 108-112.

Jiang, S. G., Chen, L., Tang, J. H., Zhao, J. H. and Zhong, S. L. (2015) Lack of association between Hsa-Mir-499 rs3746444 polymorphism and cancer risk: meta-analysis findings. Asian Pacific Journal of Cancer Prevention 16(1):339-344.

Kabirizadeh, S., Azadeh, M., Mirhosseini, M., Ghaedi, K. and Tanha, H. M. (2016) The SNP rs3746444 within mir-499a is associated with breast cancer risk in Iranian population. Journal of Cellular Immunotherapy 2(2): 95-97.

Khurana, E., Fu, Y., Chakravarty, D., Demichelis, F., Rubin, M. A. and Gerstein, M. (2016) Role of non-coding sequence variants in cancer. Nature Reviews Genetics 17(2): 93.

Machiela, M. J. and Chanock, S. J. (2015) LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics 31(21): 3555-3557.

Meng, F., Yuan, G., Zhu, X., Zhou, Y., Wang, D. and Guo, Y. (2018) Functional variants identified efficiently through an integrated transcriptome and epigenome analysis. Scientific reports 8(1): 1-13.

Mu, K., Wu, Z. Z., Yu, J. P., Guo, W., Wu, N., Wei, L. J. and Liu, J. T. (2017) Meta-analysis of the association between three microRNA polymorphisms and breast cancer susceptibility. Oncotarget 8(40): 68809.

Ong, C. T. and Corces, V. G. (2011) Enhancer function: new insights into the regulation of tissue-specific gene expression. Nature Reviews Genetics 12(4): 283-293.

Rhie, S. K., Coetzee, S. G., Noushmehr, H., Yan, C., Kim, J. M., Haiman, C. A. and Coetzee, G. A. (2013) Comprehensive functional annotation of seventy-one breast cancer risk Loci. PloS one 8(5): e63925.

Seyedmir, F., Mirzaie, K. and Bitaraf Sani, M. (2017) The Studies of Decision Tree in Estimation of Breast Cancer Risk by Using Polymorphism Nucleotide. Journal of Shahid Sadoughi University of Medical Sciences 25(4): 300-310.

Wang, L., Qian, S., Zhi, H., Zhang, Y., Wang, B. and Lu, Z. (2012) The association between hsa-miR-499

T> C polymorphism and cancer risk: a meta-analysis. Gene 508(1): 9-14.

Wang, Y., Yang, B. and Ren, X. (2012) Hsa-miR-499 polymorphism (rs3746444) and cancer risk: a meta-analysis of 17 case–control studies. Gene 509(2): 267-272.

Ward, L. D. and Kellis, M. (2012) HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic acids research 40(D1): D930-D934.

Zou, P., Zhao, L., Xu, H., Chen, P., Gu, A., Liu, N. and Lu, A. (2012) Hsa-mir-499 rs3746444 polymorphism and cancer risk: a meta-analysis. Journal of biomedical research 26(4): 253-259.

Research Article

Element Accumulations in Liver and Kidney Tissues of Some Bony Fish Species in the Southwest Caspian Sea

Masoud Sattari[1,2*], Mehdi Bibak[1], Shima Bakhshalizadeh[2], Mohammad Forouhar Vajargah[1]

[1] *Department of Fisheries, Faculty of Natural Resources, University of Guilan, Sowmeh Sara, Iran*
[2] *Department of Marine Sciences, Caspian Sea Basin Research Center, University of Guilan, Rasht, Iran*

Received 11 March 2020 Accepted 16 May 2020

Abstract

The Caspian Sea is the largest inland body of water in the world and so has both common characteristics of seas and lakes with over 153 fish species which inhabit the sea and its basin. However, little is known about the trace element (TE) contaminations (TECs) in its tissues. In the present study, 122 specimens of three fish species including *Rutilus caspius* (Roach, n=71), *Leuciscus aspius* (Asp, n=20), and *Tinca tinca* (Tench, n=31) were collected from three different fisheries regions (i.e. Astara, Anzali and Kiashahr) of the southern part of the Caspian Sea from September 2017 to June 2018. Inductively coupled plasma optical emission spectrometry (ICP-OES) was employed to measure TE levels in different fish tissues. An attempt was made to assess possible influences of habitat on element accumulation in the liver and kidney of three fish species in the southwest of the Caspian Sea basin. Some elements including Ca, K, Mg, P, S, Sc, and Sr showed different concentrations in the liver and kidney. Also their levels were significantly different between freshwater resident (Tench) and marine (Roach) species ($p < 0.05$). The differences among TECs in the liver and kidney of Roach, Asp and Tench were reduced to three components using principal component analysis (PCA). Results indicated that 83.60% of the total variability is related to TEs such as Cu, Fe, Sr, Ca, S, Na, Mg, K, and Al. The impact of habitat variability on the element accumulation was confirmed through linear chart obtained for liver and kidney (as body filtering organs) of Roach and Asp as marine residents as well as Tench as a freshwater resident. This could illustrate the borderline created by these habitats.

***Keywords*: *Rutilus caspius*, *Leuciscus aspius*, *Tinca tinca*, Trace elements, Caspian Sea**

Introduction

The trace element (TE) pollution in water resources has long been found to be a serious environmental concern (Pagano et al., 2017; Capillo et al., 2018; Chorehi et al., 2013). Aquatic organisms can accumulate TEs in their bodies via respiration, adsorption and ingestion (Zhou et al., 2001; Boran et al., 2000). TE contamination is a serious problem in the coastal regions, due to waste disposal of discharges from agriculture, industries and some urban sources (Aliko et al., 2018; Burgos-Aceves et al., 2018). TE accumulation elevated in marine ecosystems as a direct result of anthropogenic activities (Seco-Gesto et al., 2007). TEs were categorized as potentially toxic (cadmium, arsenic, lead, mercury, nickel, etc.), probably essential (vanadium, cobalt) and essential (copper, selenium, iron, manganese, zinc) (Munoz-Olivas et al., 2001). Fish is considered as a suitable indicator for long term monitoring of TE contaminations in different water resources (Fazio et al., 2014; Sattari et al., 2019). Therefore, numerous studies have been conducted on TE accumulation in different fish species (Türkmen et al., 2007).

The Caspian Sea is the world's largest inland body of water and thus has characteristics common to both seas and lakes. It is bordered by Russia (Dagestan, Kalmykia, and Astrakhan oblasts), the Republic of Azerbaijan, Iran (Guilan, Mazandaran and Golestan provinces), Turkmenistan, and Kazakhstan (Vajargah et al., 2014; Sattari et al., 2019). Hence, not only it doesn't contain fresh water, but also it is under intense pollutant threats from industrial and agricultural effluents as well as growing urbanization in the most riparian countries of the Caspian Sea (Karrari et al., 2012).

There are numerous reports regarding the heavy metal contamination in aquatic environments of Iran including: Pourang et al., (2005) on five sturgeon species in the Caspian Sea; Abtahi et al., (2005) on *Liza aurata* in the south Caspian Sea; Sadeghirad, (2007) on *Acipenser persicus* and

*Corresponding author's e-mail address: *msattari@guilan.ac.ir*

Acipenser stellatus from the Caspian Sea; De Mora et al., (2004) on coastal sediments from the Caspian Sea; Amini Ranjbar and Sotudehnia, (2005) on *Mugil auratus* of the Caspian Sea; Askary Sary and Beheshti, (2012) on *Liza abu* from the Karoun and Karkheh rivers; Beheshti, (2011) on *Liza abu* collected from the Karkheh and Karoon rivers; Ebrahimzadeh et al., (2011) on *Liza saliens* collected from the Caspian Sea; Khanipour et al., (2018) on *Silurus glanis* collected from Anzali Wetland, the southwest Caspian Sea; and Alipour and Banagar, (2018) on fish obtained from Gorgan Bay, the southeast Caspian Sea. There are also some reports on TE accumulations in *R. kutum* (Shahryari et al., 2010; Eslami et al., 2014; Mirzajani et al., 2016; Sattari et al., 2019). Eslami et al., (2011) reported the existence of TEs in muscle and liver of *Perca fluviatilis* and *Tinca tinca* in Anzali Wetland; Bibak et al., (2018) worked on heavy metal levels in sediments of the northern part of the Persian Gulf. This study aimed to determine the levels of some target trace elements (TEs) in the livers and kidneys of three fish species which were collected from the geographically different coastal regions of the Caspian Sea.

Materials and Methods

This study was conducted in three fisheries regions including Kiashahr (37° 42′ 20″ N, 49° 94′ 95″ E), Astara (38° 42′ 25″ N, 48° 86′ 87″ E), and Anzali (37° 46′ 39″ N, 49° 47′ 99″ E) along the south western coasts of the Caspian Sea. 122 specimens were collected from September, 2017 to January, 2018 from three different fish species including *Rutilus caspius* (Roach, n=71), *Leuciscus aspius* (Asp, n=20), and *Tinca tinca* (Tench, n=31) with gill net. The specimens were transported to the Fish Biology Laboratory, University of Guilan, Sowmeh Sara, Iran by a styrofoam cooler box at 4°C. Fish were washed using distilled water, dissected and pieces of muscle were dried in the oven (80°C for 18 h) (Vajargah et al., 2018b). Fish age was determined with scales during the process. To extract TEs, 0.5 g of each tissue was digested in 10 ml of 65% nitric acid in a microwave oven. Then, specimens were passed through the Whatman filter paper No. 40 and were diluted in distilled water to the required volume. An inductively coupled plasma–optical emission spectrometry (ICP-OES) (Zarazma Co. Tehran, Iran) was employed to measure trace element levels in the specimens. Instrumental detection limits for trace and major (Al, Ca, Fe, K, Mg, Mn, Na and Si) element measurements were equal to 0.02 $mgkg^{-1}$, and 0.1 mg kg^{-1}, respectively.

Statistical analyses

After examining the normality of acquired data and homogeneity of variances in the fish tissues (liver and kidney) from different habitats, the variability of TE concentrations was investigated through one-way analysis of variances (ANOVA). For heterogeneous variables, the Kruskal-Wallis test was employed, otherwise, we used Man-Whitney U test (Zar, 1996).

Principle component analysis (PCA) was used to reduce the number of variables without losing much information (Quinn and Keouch, 2002). Eigenvalues against the number of principal components and also the values of cumulative variances were provided to define the important principle components and elements. Discriminant function analysis (DFA) was employed to calculate the exact place of each fish species which were related correctly. Ward's method, as a complementary method for DFA, was employed to construct cluster dendrogram using Euclidean distance (average linkage clustering). All statistical analyses were performed using SPSS version 16.0 (significance level α=0.05, SPSS Inc., Chicago, IL, USA).

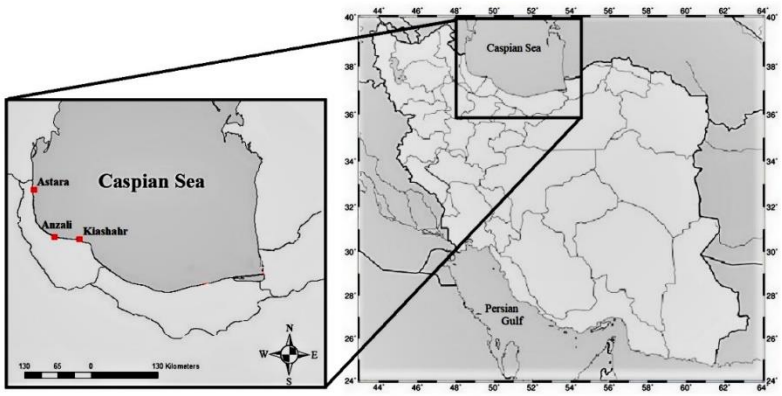

Figure 1. Map of the study area along the south western coasts of the Caspian Sea

Results

In the present study, several specimens from three different fish species were dissected and their kidney and liver tissues were examined for the presence of 36 elements including: Silver (Ag), Aluminum (Al), Arsenic (As), Barium (Ba) Beryllium (Be), Bismuth (Bi), Calcium (Ca), Cadmium (Cd), Cesium (Ce), Cobalt (Co), Chromium (Cr), Copper (Cu), Iron (Fe), Potassium (K), Lanthanum (La), Lithium (Li), Magnesium (Mg), Manganese (Mn), Molybdenum (Mo), Sodium (Na), Nickel (Ni), Phosphorus (P), lead (Pb), Rubidium (Rb), Sulfur (S), Antimony (Sb), Scandium (Sc), Silicon (Si), Tin (Sn), Strontium (Sr), Thorium (Th), Titanium (Ti), Uranium (U), Vanadium (V), Tungsten (W), Yttrium (Y) and Zinc (Zn).

Some elements such as Ca, K, Mg, P, S, Sc and Sr exhibited different concentrations between liver and kidney (*p<0.05, Table 1). Their levels were also displayed significant differences between Tench, as a freshwater resident, and Roach, as a marine species (Table 2). The variability of TECs in different edible tissues of Roach, Asp and Tench was reduced to three components using PCA (PC1= 46.32%, PC2 =22.48% and PC3=14.79%) (Figures 2-3). It was found that 83.60% of total variation is related to TEs such as Cu, Fe, Sr, Ca, S, Na, Mg, K and Al (Table 3). The three-dimensional diagram illustrated the weight of each component in PCA is shown in Figure 3 (Figure 3). The first component was mainly influenced by Na, Mg, K and S (Table 3). Ca and Sr had special contributions in PC2, while the highest values in PC3 were obtained for Cu, Fe and Al (Table 3).

Table 1. Mean of concentration±SD for different elements in the liver and kidney of the sea water fish and lagoon fish of the Iranian Caspian Sea (*p <0.05).

Elements	Elemental concentration (ppm; Median ± SD)			*p* value
	T. tinca	***R. caspius***	***L. aspius***	
Al	019 ± 0.20	0.22 ± 0.11	0.05 ± 0.05	0.46
Ca	9.06 ± 0.38	25.84 ± 9.73	97.30 ± 73.30	0.02
Cr	0.00 ± 0.00	0.02 ± 0.01	0.09 ± 0.01	0.12
Cu	0.07 ± 0.05	0.17 ± 0.12	0.00 ± 0.00	0.36
Fe	1.59 ± 0.59	4.82 ± 1.92	0.63 ± 0.25	0.17
K	38.33 ± 12.97	114.62 ± 8.01	115.30 ± 67.60	0.03
Mg	4.03 ± 0.81	15.06 ± 2.33	17.05 ± 5.25	0.02
Na	20.57 ± 5.78	56.16 ± 6.25	48.50 ± 19.70	0.58
P	36.27 ± 11.60	117.00 ± 13.94	220.60 ± 119.9	0.02
S	29.73 ± 10.18	99.18 ± 8.56	92.70 ± 52.00	0.03
Sb	0.08 ± 0.08	0.04 ± 0.02	0.09 ± 0.09	0.59
Si	0.00 ± 0.00	0.30 ± 0.06	0.26 ± 0.06	0.02
Sr	0.05 ± 0.02	0.26 ± 0.12	0.72 ± 0.63	0.02
Zn	0.70 ± 0.12	2.08 ± 0.28	1.20 ± 0.30	0.08

Table 2. Pair-wise comparisons (Mann–Whitney U-test) of the significant elements among the fish belongs to sea water and lagoon of the Iranian Caspian Sea.

Elements	Tench-roach			Tench-asp			Roach-asp		
	M-W	Z	*p*	M-W	Z	*P*	M-W	Z	*p*
Ca	0.00	-2.24	<0.05	0.00	-1.73	>0.05	1.00	-1.55	>0.05
k	0.00	-2.24	<0.05	1.00	-1.16	>0.05	5.00	0.00	>0.05
Mg	3.00	-2.24	<0.05	1.00	-1.73	>0.05	0.00	-0.78	>0.05
P	0.00	-2.24	<0.05	0.00	-1.73	>0.05	3.00	-0.78	>0.05
S	0.00	-2.24	<0.05	0.00	1.00	>0.05	1.00	5.00	>0.05
Sr	0.00	-2.24	<0.05	0.00	-1.73	>0.05	4.50	-0.20	>0.05
Si	0.00	-2.29	<0.05	0.00	-1.94	>0.05	4.50	-0.20	>0.05

M-W and Z are grouping variable scores for Mann–Whitney U and Kruskal–Wallis tests (*p <0.05).

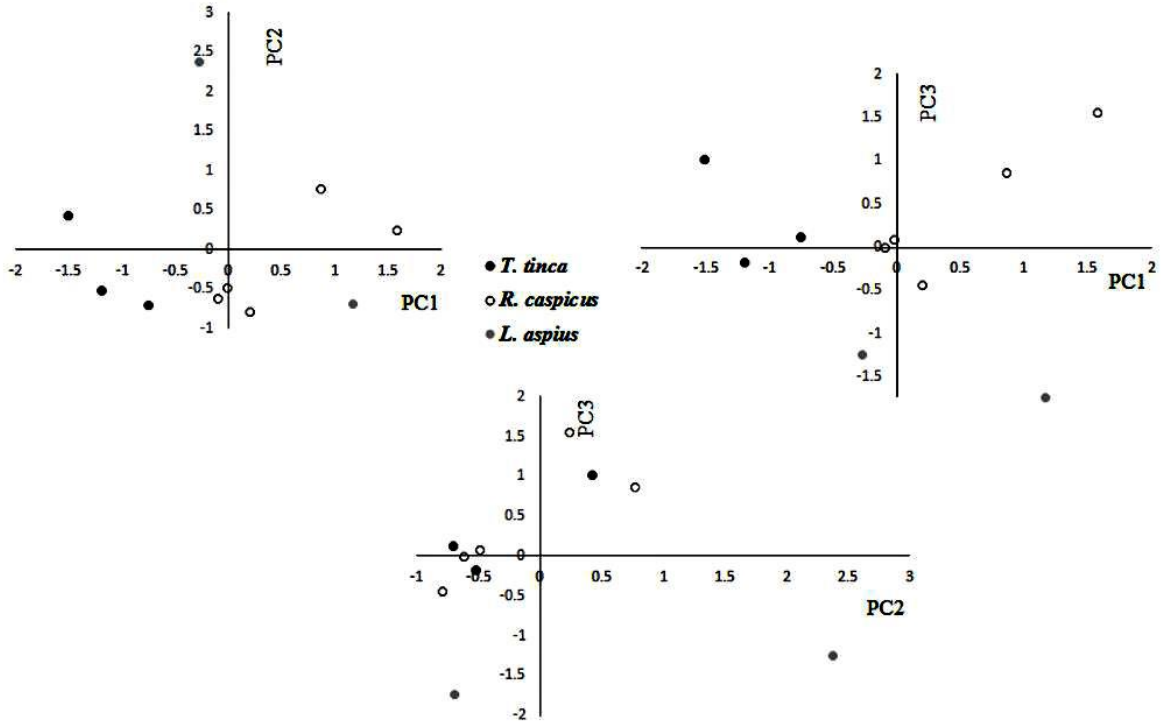

Figure 2. Principal component analysis (PCA) of elemental concentrations of the fish liver and kidney between habitats in the coastal water of the Iranian Caspian Sea. Scatter plots demonstrate individual fish scores for PC1 vs. PC2, PC1 vs. PC3, and PC2 vs. PC3 which together explain 83.60% of the total variance, this graph obtained by SPSS version 16.0.

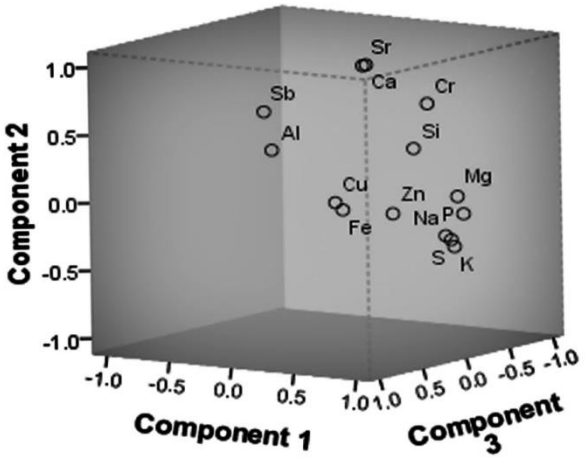

Figure 3. Characteristic load for PC1, PC2 and PC3 obtained by multi-elemental principal components analysis (PCA) for the elemental concentrations of the fish liver and kidney between habitats in the south Caspian Sea water, this graph obtained by SPSS version 16.0.

Table 3. Characteristic load for PC1, PC2 and PC3 obtained by principal component (PCA) analysis for elemental concentrations of the fish liver and kidney between habitats (sea water and lagoon) in the water of Iranian Caspian Sea

Elemental variables	*PC1	PC2	PC3
Al	.012	.418	.629
Ca	.087	.903	-.361
Cr	.576	.661	-.360
Cu	.544	.084	.666
Fe	.615	.039	.683

K	.906	-.337	-.179
Mg	.965	.044	-.138
Na	.933	-.233	-.041
P	.777	-.152	-.473
S	.952	-.267	-.078
Sb	-.287	.622	.286
Si	.844	.437	.188
Sr	.174	.931	-.199
Zn	.718	-.046	.256

***** Principle Components

The matrix composed of element concentrations in liver and kidney tissues of Roach, Asp, and Tench was described with two discriminant components. These experiments successfully discriminate the two investigated habitats (Wilk's Lambda=0.001, X^2=27.63, df=14 and p<0.05, Figure 4). Cluster analysis, as a complementary method, divided the fish into two sub-groups. No variation was found for element concentrations in Roach and Asp, while Tench was placed in a distinct subgroup (Figure 5).

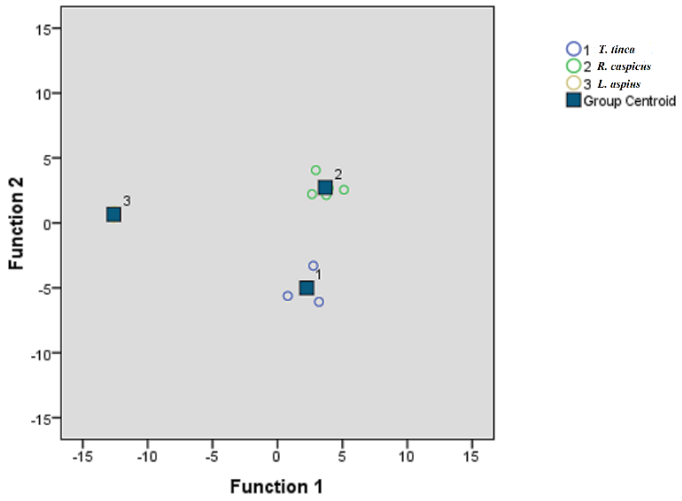

Figure 4. Plot of discriminant functions 1 and 2 for the elemental concentrations of the fish liver and kidney between habitats in the coastal water of the Iranian Caspian Sea, this graph obtained by SPSS version 16.0.

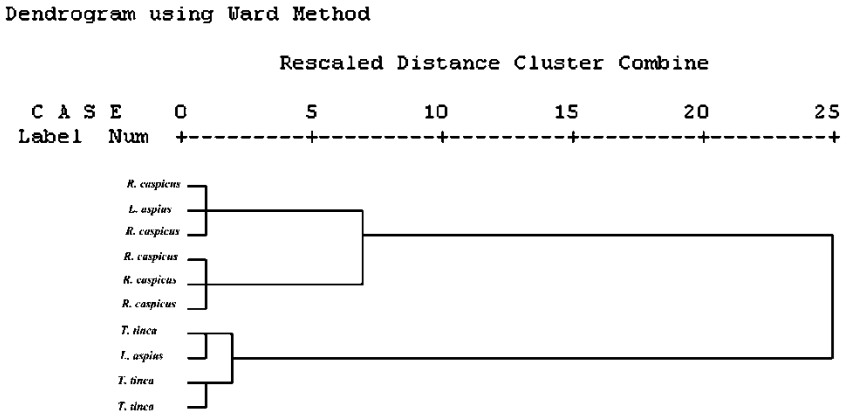

Figure 5. Dendrogram derived from cluster analysis of the elemental concentrations of the fish liver and kidney between habitats in the coastal water of the Iranian Caspian Sea, Cluster analysis, as a complementary method, divided the fish into two sub-groups. No variation was found for element concentrations in Roach and Asp, while Tench was placed in a distinct subgroup, this graph obtained by SPSS version 16.0.

Discussion

In the present study, an attempt was made to assess possible influences of habitat on the elements accumulation in the liver and kidney tissues of some fish species in the southwest of the Caspian Sea basin. Linear charts of element accumulations in the body filtering organs of Roach, Asp, and Tench exhibited their enough variability based on the habitats and also illustrated the borderline created by these habitats. The consequences of different elements bioaccumulation on fish tissues depend on sex, maturation stage, size, tissue type, habitat and fish diet (Azevedo et al., 2009). Previous studies revealed that element bioaccumulation in various fish tissues take place at different levels; but, in short time periods, filtering organs such as liver exhibit higher levels of these elements (Afonso et al., 2017, Alamdar et al., 2017; Salgado-Ramírez et al., 2017). This is while, gills and gut are the first organs receiving these elements (Tiphaine et al., 2018). Therefore, liver could be considered as the main organ for element aggregation monitoring studies (Salgado-Ramírez et al., 2017)

Since metal elements find their way to the aquatic environment and deposit in sediments, the demersal and benthivorous fish species are more susceptible to element bioaccumulation than planktivorous fish (Trevizani et al., 2019).

Little is known about Cu accumulation in fish species. However, it has been found in the higher than normal levels in the food chain which is distinguishable from low-level elements. It seems that raising in Metallothionein levels could be considered as an exact indicator of Cu existence in the ambient environment (Marijić and Raspor, 2007). This is reduced upon the migration of the fish from saltwater to freshwater (Ohji et al., 2007). It is also true for Sr which is found in higher amounts in saltwater in comparison to freshwater. So that, Strontium levels are higher in fish tissues with long residence times in sea water, regardless of the fish diet.

Meanwhile, there is a positive correlation between Sr and Ca. So that, its concentration is raised by low temperature and high salinity (Walther and Thorrold, 2006), instead of the fish diet in Asp tissues. Overall, the present study provides some basic information about elements bioaccumulation in the fish filtering tissues from different ecosystems with various salinity levels. These data could be applicable in determining the focal points of contaminations.

Acknowledgments

This study was financially supported by The Caspian Basin Research Center, University of Guilan, Rasht, Iran (grant ID: 21195170).

References

Abtahi B. and Sabbagh kashani A. (2005) Assessing Pb, Ni and Zn accumulation in the tissues of Liza aurata in the south Caspian Sea. Iranian Scientific Fisheries Journal 14: 65-78.

Afonso A., Gutiérrez A. J., Lozano G., González-Weller D., Rubio C., Caballero J. M., Hardisson A. and Revert C. (2017) Determination of toxic metals, trace and essentials, and macronutrients in Sarpa salpa and Chelon labrosus: risk assessment for the consumers. Environmental Science and Pollution Research 24(11): 10557-10569.

Alamdar A., Eqani S. A. M. A. S., Hanif N. Ali S. M., Fasola M., Bokhari H., Katsoyiannis I. A. and Shen H. (2017) Human exposure to trace metals and arsenic via consumption of fish from river Chenab, Pakistan and associated health risks. Chemosphere 168: 1004-1012.

Aliko V., Qirjo M. and Sula E. (2018) Antioxidant defense system, immune response and erythron profile modulation in Gold fish, Carassius auratus, after acute manganese treatment" Fish Shellfish Immunology 76:101–109.

Alipour H. and Banagar G. h. R. (2018) Health risk assessment of selected heavy metals in some edible fishes from Gorgan Bay, Iran. Iranian Journal of Fisheries Sciences 17(1): 21-34.

Amini Ranjbar G. h. and Sotudehnia F. (2005) Investigation of heavy metals accumulation in muscle tissue of Mugil auratus in relation to standard length, weight, age and sex. Iranian Scientific Fisheries Journal 14: 1-18. (In Persian).

Askary Sary A. and Beheshti M. (2012) Cadmium, iron, lead and mercury bioaccumulation in Abu Mullet, Liza abu, different tissues from Karun and Karkheh Rivers, Khozestan, Iran. Bulletin of Environmental Contamination and Toxicology 88: 158-161.

Azevedo J., Fernandez W., Farias L., Fávaro D. and Braga E. (2009) Use of Cathorops spixii as bioindicator of pollution of trace metals in the Santos Bay, Brazil. Ecotoxicology 18(5): 577-586.

Beheshti M. (2011) Comparative study of concentration of heavy metals (Cu, Fe, Zn, Mn) in muscle, liver and gill organ of fish (Liza abu) in the Karoon and Karkheh rivers in Khuzestan province." M.Sc. Thesis. Islamic Azad University, Science and Research, Ahwaz. (In Persian).

Bibak M., Sattari M., Agharokh A., Tahmasebi S. and Imanpour Namin J. (2018) Assessing some heavy metals pollutions in sediments of the northern Persian Gulf (Bushehr province). Environmental Health Engineering and Management Journal 2018: 5(3): 175–179.

Boran M., Karacam H., Celikkale M. S. and Kose S. (2000) Levels of heavy metals in blue whiting caught from the eastern Black Sea area of Turkey. Toxicological and Environmental Chemistry 2000: 75, 67-73.

Burgos-Aceves M. A., Cohen A., and Smith Y. (2018) MicroRNAs and their role on fish oxidative stress during xenobiotic environmental exposures. Ecotoxicology and Environmental Safety. 148: 995-1000.

Capillo G., Silvestro S. and Sanfilippo M. (2018) Assessment of electrolytes and metals profile of the Faro Lake (Capo Peloro Lagoon, Sicily, Italy) and its impact on *Mytilus galloprovincialis*. Chemestry & Biodiversity.

Chorehi M. M., Ghaffari H., Hossaini S. A., Niazie E. H. N., Vajargah M. F. and Hedayati A. (2013) Acute toxicity of Diazinon to the Caspian vimba, *Vimba vimba persa* (Cypriniformes: Cyprinidae). International Journal of Aquatic Biology 1(6): 254-257.

De Mora S., Sheikholeslami M. R., Wyse E. and Azemard S. (2004) An assessment of metal contamination in coastal sediments of the Caspian Sea. Marine Pollution Bulletin 48: 61-77.

Ebrahimzadeh M. A., Eslami S. and Nabavi S. F. (2011) Determination of trace element level in different tissues of the leaping mullet (Lizasaliens, Mugilidae) collected from Caspian Sea. Biological Trace Element Research 144: 804-811.

Eslami S., Moghaddam A. H. and Jafari N. (2011) Trace element level in different tissues of *Rutilus frisii kutum* collected from Tajan River, Iran. Biological Trace Element Research 143: 965-973.

Eslami V., Sattari M. and Namin J.I. (2014) Concentration of heavy metals (Pb, Cd) in muscle and liver of Perca fluviatilis and *Tinca tinca* in Anzali Wetland, southwest the Caspian Sea. International Journal of Aquatic Biology 2(6): 319-324.

Fazio F., Piccione G. and Tribulato K. (2014) Bioaccumulation of heavy metals in blood and tissue of striped mullet in two Italian lakes. Journal of Aquatic Animal Health 26: 278–284.

Khanipour A. A., Ahmadi M. and Seifzadeh M. (2018) Study on bioaccumulation of heavy metals (cadmium, nickel, zinc and lead) in the muscle of wels catfish (*Silurus glanis*) in the Anzali Wetland. Iranian Journal of Fisheries Sciences 17(1): 244-250.

Marijić V. F. and Raspor B. (2007) Metallothionein in intestine of red mullet, *Mullus barbatus* as a biomarker of copper exposure in the coastal marine areas. Marine pollution bulletin 54(7): 935-940.

Munoz-Olivas R. and Camara C. (2001) Speciation related to human health, In: L, Ebdon, L, Pitts, R, Cornelis, H, Crews, OFX, Donad, P, Quevauviller, (Eds.), Trace Element Speciation for Environment, Food and Health. The Royal Society of Chemistry 331-353.

Ohji M., Arai T. and Miyazaki N. (2007) Comparison of organotin accumulation in the masu salmon *Oncorhynchus masou* accompanying migratory histories. Estuarine, Coastal and Shelf Science 72(4): 721-731.

Ohji M., Harino H. and Arai T. (2006) Differences in organotin accumulation among ecological migratory types of the Japanese eel Anguilla japonica. Estuarine, Coastal and Shelf Science 69(1-2): 270-290.

Pagano M., Porcino C. and Briglia M. (2017) The influence of exposure of cadmium chloride and zinc chloride on haemolymph and digestive gland cells from *Mytilus galloprovincialis*. International Journal of Environmental Research 11(2): 207-216.

Pourang N. and Dennis J. H. (2005) Distribution of trace elements in tissues of two shrimp species from the PersianGulf and roles of metallothionein in their redistribution. Environment International 31(3):325–341.

Sadeghirad M. (2007) Accumulation of trace elements in Acipenser persicus tissues in relation to feeding habits and mode of absorption of these pollutants. Iranian fisheries research organization. Internatinoal sturgeon research institut. 42P.

Salgado-Ramírez C. A., Mansilla-Rivera I. and Rodríguez-Sierra C. J. (2017) Comparison of trace metals in different fish tissues of Scomberomorus spp.("sierra") and *Lutjanus synagris* ("arrayado") from Jobos Bay and La Parguera coastal areas in southern Puerto Rico. Regional studies in marine science 13: 1-11.

Sattari M., Imanpour J., Bibak M., Forouhar Vajargah M. and Khosravi A. (2019) Investigation of metal element concentrations in tissue of Rutilus frisii in the Southwest Caspian Sea. Iranian Scientific Fisheries Journal 28(3): pp.149-161.

Sattari M., Namin J. I., Bibak M., Vajargah M.F., Bakhshalizadeh S. and Faggio, C. (2019) Determination of Trace Element Accumulation in Gonads of Rutilus kutum (Kamensky, 1901) from the South Caspian Sea Trace Element Contaminations in Gonads. Proceedings of the National Academy of Sciences, India Section B: Biological Sciences

Sattari M., Namin J. I., Bibak M., Vajargah M. F., Faggio C. and Haddad M. S. (2019) Trace and macro elements bioaccumulation in the muscle and liver tissues of *alburnus chalcoides* from the south Caspian Sea and potential human health risk assessment. Journal of Energy, Environmental & Chemical Engineering 4(1):13.

Sattari M., Namin J. I., Bibak M., Vajargah M.F., Hedayati A., Khosravi A. and Mazareiy M.H. (2019) Morphological comparison of western and eastern populations of Caspian kutum, Rutilus kutum (Kamensky, 1901)(Cyprinidae) in the southern Caspian Sea. International Journal of Aquatic Biology 6(4): 242-247.

Seco-Gesto E. M., Moreda-Piñeiro A. and BermejoBarrera A. (2007) Multi-element determination in raft mussels by fast microwave assisted acid leaching and inductively coupled plasma-optical emission spectrometry, Talanta. 72: 1178.

Shahryari A., Golfirozy K., Noshin S. (2010) Muscular concentration of cadmium and lead in carp, mullet and kutum of the Gorgan Bay, Caspian Sea. Iranian Journal of Fisheries Sciences 19(2): 95-100.

Tiphaine M., Pierre C., Tiphaine C., Paco B., Christophe B.P., Sandrine B., Emmanuelle R. and Marc B. (2018) Trace metal concentrations in the muscle of seven marine species: Comparison between the Gulf of Lions (North-West Mediterranean Sea) and the Bay of Biscay (North-East Atlantic Ocean). Marine pollution bulletin 135: 9-16.

Trevizani T. H., Domit C., Vedolin M. C., Angeli J. L. F. and Figueira R. C. L. (2019) Assessment of metal contamination in fish from estuaries of southern and southeastern Brazil. Environmental monitoring and assessment 191(5): 308.

Türkmen M., Türkmen A. and Tepe Y. (2007) Metal contaminations in five fish species from Black, Marmara, Aegean and Mediterranean seas, Turkey. Journal of the Chilean Chemical Society 52: 4, 1314-1318.

Tzeng W., Shiao J., and Iizuka Y. (2002) Use of otolith Sr: Ca ratios to study the riverine migratory behaviors of Japanese eel Anguilla japonica. Marine Ecology Progress Series 245: 213-221.

Vajargah M. F., Hedayati A., Yalsuyi A. M., Abarghoei S., Gerami M. H. and Farsani, H. G. (2014) Acute toxicity of butachlor to caspian kutum (*Rutilus frisii Kutum* Kamensky, 1991). Journal of Environmental Treatment Techniques 2(4): 155-157.

Walther B. D. and Thorrold S. R. (2006) Water, not food, contributes the majority of strontium and barium deposited in the otoliths of a marine fish. Marine Ecology Progress Series 311: 125-130.

Zhou J. L., Salavador S. M. and Liu Y. P. (2001) Heavy metals in the tissues of common dolphins (*Delphinus delphis*) stranded on the Portuguese coast. Science of the Total Environment 273: 61.

Scientific Reports

Association of *TOMM40*, *CHAT* and *SORL1* Polymorphisms with the Alzheimer's Disease in the Turkish-speaking Azeri population in Northwest of Iran

Nazanin Mardokh Rohani[1], Mohammad Khalaj-Kondori[1*], Majid Khodayi-Shahrak[1], Mahnaz Talebi[2], Mohammad Ali Hoseinpur Feizi[1]

[1]*Department of Animal Biology, Faculty of Natural Sciences, University of Tabriz, Tabriz, Iran*
[2]*Neurosciences Research Center, Tabriz University of Medical Sciences, Tabriz, Iran*

Received 14 June 2020 Accepted 9 September 2020

Abstract

Recent genome-wide association studies have introduced several genetic variants which contribute to the late-onset Alzheimer's disease (LOAD). Polymorphisms of *CHAT*, *TOMM40*, and *SORL1* genes have been reported to be associated with the LOAD phenotype. This study was endeavored to evaluate the association of the *CHAT* rs3810950, *TOMM40* rs1160985 and *SORL1* rs11218304 polymorphisms with the LOAD in the Turkish-speaking Azeri population of northwest Iran. In a case-control study, we included 174 cases: 88 cases with LOAD diagnosis and 86 healthy individuals. Peripheral blood samples were collected and the genomic DNA of all participants were extracted. Genotyping was carried out by the polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) method. We did not observe any significant association between the *CHAT* rs3810950 and *SORL1* rs11218304 alleles with the LOAD. However, both the *TOMM40* rs1160985 minor allele T and TT genotype showed significant negative associations with the LOAD. Hence, the *TOMM40* rs1160985 polymorphism could be considered as a protective genetic factor against the LOAD in the Turkish-speaking Azeri population of northwest Iran.

***Keywords*:** Alzheimer's disease, *SORL1*, *CHAT*, *TOMM40*, Genome-wide association study

Introduction

Dementia syndrome due to the Alzheimer's disease (AD) is one of the most expensive chronic diseases with powerful threat (Belmonte et al., 2015). According to the 2019 Alzheimer's Disease Facts and Figures, of 5.6 million persons aged 65 and older with Alzheimer's in the United States, 3.5 million are women and 2.1 million are men (Association, 2019). Considering the ever-increasing feature of the disease, it is estimated that the number of individuals with AD will be more than 15 million in 2060 (Brookmeyer et al., 2018). These warns highlighted the urgent need for the development of new diagnostics and therapeutics; ranging from biomarker discovery (Fotuhi et al., 2019; Yanfang Zhao, 2019) to *in vivo* reprogramming of the terminally differentiated cells (Yavarpour-Bali et al., 2020).

AD occurs in familial and non-familial forms (early vs. late age-onset, respectively). Both genes and environment are responsible for the appearance of the non-familial sporadic late-onset AD (LOAD), as a complex disorder (Bertram et al., 2010). Genetic factors are estimated to play a role at least in 80% of AD cases (Gatz et al., 2006; Tanzi, 2012). In addition, at least up to age of 80, having a family history of AD increases the risk of developing disease up to 4 to 10 folds (Honea et al., 2012). Recent case-control and genome-wide association studies (GWAS) have partly revealed the genetic origin of the LOAD and highlighted its complex nature (Liu et al., 2016; Ortega-Rojas et al., 2016; Talebi et al., 2020; Yuan et al., 2016). Based on these reports, *CHAT*, *SORL1*, and *TOMM40* are important genes in the LOAD pathogenesis.

The gene encoding for TOMM40 (Translocase of outer mitochondrial membrane 40 homolog) is located on the chromosome19, closely next to the gene which is encode for Apolipoprotein E (ApoE). So, it has a strong linkage disequilibrium (LD) with it (Lyall et al., 2014). TOMM40, the central and key subunit of the translocase of the outer mitochondrial membrane, is essential for protein import into the mitochondria. Genetic variations in or close to the *TOMM40* gene affect the role of the TOMM40,

[*]Corresponding author's e-mail address: *khalaj@tabrizu.ac.ir*

thereby, causing mitochondrial dysfunction (Petschner et al., 2018). Involvement of *TOMM40* in the LOAD pathogenesis has been proposed by several researchers (Petschner et al., 2018; Willette et al., 2017), however, its role in LOAD pathogenesis was controversial (Yu et al., 2007). Yu, et al. reported a significant LD between *TOMM40* and *APOE* in the Caucasians. The involvement of *TOMM40* in LOAD was further supported by some later studies (Jiao et al., 2015; Omoumi et al., 2014; Ortega-Rojas et al., 2016; Roses et al., 2016). Interestingly, two studies on the Chinese and Columbian populations reported that *TOMM40* rs1160985 might be useful for early diagnosis of the LOAD (Jiao et al., 2015; Ortega-Rojas et al., 2016). *CHAT* (Choline O-acetyltransferase) encodes an enzyme which is crucial for the synthesis of acetylcholine, one of the main neurotransmitters in the brain. The choline acetyltransferase (*ChAT*) activity seems to be associated with the severity of dementia (Gao et al., 2016), and its polymorphisms are known to be also related with the LOAD (Thangnipon et al., 2016; Yu et al., 2015). Conversely, others did not find any significant association between *CHAT* polymorphisms and LOAD (Cook et al., 2005).

Sortilin related receptor 1 (*SORL1*) functions as a neural sorting factor (Felsky et al., 2014). It transfers the amyloid precursor protein (APP) to the recycling pathway and hinders the beta amyloid formation in the brain (Rogaeva et al., 2007). Some studies have reported the association of *SORL1* with the LOAD (Rosenberg et al., 2016); but, others didn't approve this association (Rogaeva et al., 2007) or reported inconsistent findings (Reynolds et al., 2013). Controversial findings also were recorded for *SORL1* rs11218304. Rogaeva et al. reported no relationship between this variant with the LOAD (Rogaeva et al., 2007), while others indicated significant associations between rs11218304 and LOAD (Louwersheimer et al., 2015; Shao et al., 2017).

This study aimed to investigate genotypes and alleles frequencies of the polymorphisms rs3810950 (*CHAT*), rs11218304 (*SORL1*), and rs1160985 (*TOMM40*) in a population from northwest of Iran and evaluate their associations with the late-onset Alzheimer's disease.

Materials and Methods

Participants

In the present case-control study, 88 patients with LOAD (53 women, 35 men) and 86 healthy voluntaries (53 women, 33 men) from the Turkish-speaking Azeri population of northwest Iran were included. All subjects were older than 65 years. The case and control groups, as far as possible, were matched for different parameters such as age and sex. All of the subjects were evaluated by a neuroscience specialist in the Clinic of the Imam Reza Medical Research Center, Tabriz, Iran. Subjects were diagnosed based on the National Institute of Neurological and Communicative Disorders (NINCDS) and Stroke and the Alzheimer's Disease and Related Disorders Association (ADRDA) criteria (Dubois et al., 2007). All cases were assessed using physical examinations and neuropsychological tests. Furthermore, the Mini-Mental State Examination (MMSE) was carried out to evaluate any cognitive deficit in both groups. The study protocol was approved by the Clinical Research Ethics Committee of Tabriz University of Medical Sciences and written informed consent was obtained from all individuals in accordance with the approved guidelines from the Neurology Department at Imam Reza Hospital. Participants with a family history of AD and other neurological illnesses such as hypothyroidism, alcoholism, hepatic lesions, spasticity, subdural hematoma, traumatic brain injury, encephalitis, frontal lobe dementia, and Lewy body dementia were excluded from the study. Participants with no memory complaint or cognitive dysfunctions and MMSE score more than 27 were defined as normal cases.

DNA preparation and genotyping

Genomic DNA were extracted from peripheral blood lymphocytes using the salting out DNA extraction method (Miller et al., 1988). The Single-nucleotide polymorphisms (SNPs) in *TOMM40* (rs1160985), *CHAT* (rs3810950), and *SORL1* (rs11218304) genes were genotyped by polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) analysis. Primer sequences and size of their amplicons are shown in the Table 1. The PCR reactions were done in a final volume of 20 µl (1 µl genomic DNA, 0.75 µl dNTPs 10 mM (Fermentas, Life Sciences), 1 µl of each of the forward and reverse primers (Metabion), 2 µl of 10× buffer, 0.5 µl MgCl2 50 mM and 1U of Taq polymerase (Sinacolon)). The optimized PCR condition was as follows: initial denaturation (95°C, 5 min), followed by 35 cycles of 95°C for 30s, 60°C for 30s, and 72°C for 40s. It was followed by a final extension (72 °C, 5 min). Then, PCR products were digested by specific restriction enzymes. In addition, 10% of the total volume of PCR products were randomly sequenced to confirm the results of RFLP

analysis.

Table 1. Primer sequences applied during PCR experiments in addition to their amplicons size.

Gene (SNP)	Length (bp)	Primer sequence	Direction
TOMM40 (rs1160985)	345bp	5′-CAAAGTGAATCCATCTCCATCC-3′	Forward
		5'-CAAGGGCAGAATCCAAGC-3'	Reverse
CHAT (rs3810950)	483bp	5'- GTTGATGCTTCCCACTTCTTG -3′	Forward
		5'-GTAGGAATTCAGCCCCACC-3'	Reverse
SORL1 (rs11218304)	386bp	5'-TCCCTCCTGTCCCGACTTC -3′	Forward
		5'-CGCATACAAGCACGCATAAG-3'	Reverse

SNP: Single-nucleotide polymorphism; bp: base pair

Genotyping of *CHAT* rs3810950

The resulting 483 bp PCR products of *CHAT* were digested with 1U of *ApeK1* (Ferments, Life Sciences) for 16h at 37°C. Final preparations were electrophoresed on agarose gel (2%) in order to identify the genotypes of each person. Samples prepared from homozygous (GG) and heterozygous (GA) genotypes were contained 2 (100 and 383 bp) and 3 (483, 100, and 383 bp) fragments, respectively, whereas a single band with 483 bp length was obtained for genotype AA.

Genotyping of *TOMM40* rs1160985

The resulting 345 bp PCR product of *TOMM40* gene was digested with 1U of *Acc1* (Ferments, Life Sciences) at 37°C for 16h. Following digestion, genotypes of the people were determined using 2% agarose gel electrophoresis. The homozygote TT and heterozygous CT genotypes contained 2 fragments (100 and 245 bp) and 3 fragments (345, 100, and 245 bp), respectively; whereas CC genotype showed a band of 345 bp.

Genotyping of *SORL1* rs11218304

The resulting 386 bp PCR product of *SORL1* gene was digested with 1U of *ApeK1* (Ferments, Life Sciences) for 16h at 37°C. Digestion products were electrophoresed on 2% agarose gel and genotypes of the people were determined. The homozygote AA and heterozygous AG contained 2 fragments (142 and 244 bp) and 3 fragments (386, 142, and 244 bp), respectively; while GG genotype showed a band of 386 bp.

Statistical analyses

The SPSS software version 21.0 (IBM SPSS, Armonk, NY, USA) was utilized for statistical analyses. The Hardy-Weinberg equilibrium (HWE) was assessed using a goodness-of-fit $\chi 2$ test. Allelic and genotypic frequencies were compared among examined groups using the Student's t-test and Odds ratio (OR) of each genotype was assessed with confidence interval (CI) 95%. *P* value ≤0.05 was considered as statistically significant.

Results

In this case-control study, 88 LOAD patients and 86 healthy individuals were enrolled. Table 2 represents demographic data of the LOAD and healthy subjects. There were no significant differences between LOAD and control groups regarding age, sex, and educational levels ($p>0.05$). Moreover, allele and genotype frequencies were calculated for *CHAT* rs3810950, *TOMM40* rs1160985, and *SORL1* rs11218304 gene polymorphisms in LOAD and control cases. The Chi-square Test revealed that the study population was in Hardy-Weinberg equilibrium for these loci.

Allele and genotype distributions of rs3810950 (*CHAT*) polymorphism

The frequency of minor allele A of *CHAT* rs3810950 polymorphism was 36% in the LOAD group and 46% in the control group; while the frequency of allele G was 64% in the LOAD and 54% in the control group. The frequencies of AA, AG, and GG genotypes were calculated as 9%, 53%, and 38% in the LOAD group, respectively. However, the frequencies of these genotypes were equal to 10%, 70%, and 19% in the control group. Statistical analysis (Table 3) revealed that the frequencies of genotype GG were significantly different between the LOAD and control groups ($p= 0.002$, OR=2.49, 95% CI=1.25-5.03).

Table 2. Sociodemographic characteristics of LOAD patients and healthy controls.

Variables	LOAD (n=88) (%)	Control (n=86) (%)	*p* value
Gender			
Female	53 (60.2)	53 (61.6)	0.85
Male	35 (39.8)	33 (38.4)	
Age (mean±SD)	71.84±6.51	71.22±5092	0.57
Education			
Illiterate (%)	54 (61.33)	50 (58.1)	
Primary (%)	26 (29.54)	31 (36.05)	
Diploma (%)	3 (3.40)	4 (4.6)	
College (%)	1 (1.13)	1 (1.2)	
MMSE (mean±SD)	19.33±5.0002	27.30±0.543	

SD: standard deviation; MMSE: mini-mental status score; AD: Alzheimer's disease; n: number.

Table 3. Allele and genotype distributions of the *CHAT* rs3810950, *TOMM40* rs1160985 and *SORL1* rs11218304 polymorphisms in the LOAD and healthy control groups.

Polymorphism	**Alleles/ Genotypes**	**LOAD (n=88)**	**Control (n=86)**	**Total population**	**HWE**	**Odds ratio (95% CI)**	***p* value**
		n (%)	**n (%)**	**n (%)**			
CHAT rs3810950	A	63 (36)	79 (46)	142 (41)		0.660 (0.375-1.164)	0.098
	G	113 (64)	93 (54)	206 (59)		1.514 (0.859-2.670)	0.098
	AA	8 (9)	9 (10)	17 (9.8)	0.00017	0.890 (0.345-2.294)	0.50
	AG	47 (53)	61 (71)	108 (62.06)		0.461(0.257-0.826)	0.007
	GG	33 (38)	16 (19)	49 (28.16)		2.49 (1.25-5.03)	0.002
TOMM40 rs1160985	T	49 (28)	98 (57)	147 (42)		0.292 (0.163-0.529)	0.000
	C	127 (72)	74 (43)	201 (58)		3.409 (1.891-6.145)	0.000
	TT	8 (9)	33 (38)	41 (23.56)	0.0019	0.161 (0.073-0.357)	0.000
	TC	33 (38)	32 (37)	65 (37.36)		1.044 (0.589-1.850)	0.50
	CC	47 (53)	21 (24)	68 (39.08)		3.571 (1.952-6.533)	0.000
SORL1 rs11218304	A	126 (72)	129 (75)	255 (73)		0.857 (.457-1.607)	0.374
	G	50 (28)	43 (25)	93 (27)		1.167 (.422-2.188)	0.374
	AA	43 (49)	43 (50)	86(49.42)	0.00084	.961(.552-1.673)	0.500
	AG	40 (45)	43 (50)	83(47.70)		.818(.469-1.426)	0.286
	GG	5(6)	0	5(2.87)		2.053 (.935-2.053)	0.030

LOAD: late-onset Alzheimer's disease; *SORL1*: Sortilin related receptor 1; *CHAT*: choline O-acetyltransferase; *TOMM40*: translocase of outer mitochondrial membrane 40 homolog; HWE: Hardy-Weinberg equilibrium; n: number; CI: confidence interval.

Allele and genotype distributions of rs1160985 (*TOMM40*) polymorphism

The frequency of minor allele T of rs1160985 polymorphism was 28% in the LOAD group and 57% in the control group while the allele C frequency was 72% in the LOAD group and 43% in the control group. Distribution of TT, TC, and CC genotypes (Table 3) for this polymorphism in the case group was 9%, 38%, and 53%, and in healthy individuals was 38%, 37%, and 24%, respectively. Statistical analysis revealed significant differences for allele C (p=0.000, OR=3.429, 95% CI=1.83-6.47) and CC genotype (p=0.000, OR=3.550, 95% CI=1.865-6.795) frequencies in case and control groups.

Allele and genotype distributions of rs11218304 (*SORL1*) polymorphism

The frequency of minor allele G of rs11218304 polymorphism was 28% in the LOAD group and 25% in the control group, while the allele A frequency was calculated as 72% and 75% in the

LOAD and control groups, respectively (Table 3). Genotype frequencies of AA, AG, and GG for rs11218304 were calculated as 49%, 45%, and 6% in the LOAD group, respectively. They were equal to 50%, 50%, and 0% in the control group. There was a significant difference between the frequencies of genotype GG in the case and control groups (p=0.030, OR=2.053, 95% CI=0.935-2.053).

Discussion

Recent meta-analyses, reviews, and genome-wide association studies have reported that the genetic variants in *TOMM40*, *CHAT*, and *SORL1* are in association with the LOAD (Campion et al., 2019; Grupe et al., 2007). In the present study, we evaluated the association of *CHAT* rs3810950, *TOMM40* rs1160985 and *SORL1* rs11218304 polymorphisms with the LOAD in the Turkish-speaking Azeri population of northwest Iran. In the case of *CHAT* rs3810950 polymorphism, the minor allele A frequency was 0.41 which is higher than all minor allele frequencies (MAFs) reported in the dbSNP (https://www.ncbi.nlm.nih.gov/snp/). However, its frequency was not significantly different between LOAD and control groups (p=0.891), which demonstrated the lack of association between the allele A and LOAD in the study population. It was while, the comparison of the genotype frequencies between the LOAD and control groups revealed a significant difference for the GG genotype (p=0.002, OR=2.49, 95% CI=1.25-5.03). These results are consistent with the results from a previous study which reported the lack of relationship between the rs3810950 (*CHAT*) polymorphism and the LOAD risk in Caucasian cohort (UK)(Cook et al., 2005). In contrast, another study performed on the Korean population by Lee et al. showed that individuals carrying the AA genotype had a significantly earlier onset of the LOAD (Lee et al., 2011). Furthermore, a meta-analysis showed that rs3810950 of *CHAT* is associated with the LOAD susceptibility (Gao et al., 2016; Yuan et al., 2016).

Moreover, we evaluated the association of *TOMM40* rs1160985 with the LOAD condition. The frequency of minor allele T in the whole study population was calculated as 0.42 which was higher than the Vietnamese people and lower than all other populations which were reported in the dbSNP (https://www.ncbi.nlm.nih.gov/snp/). Differences between minor allele T frequencies among the LOAD (0.28) and control (0.57) groups were statistically significant (p=0.000; OR=0.292; CI: 0.163-0.529); implying its negative association with the LOAD in the examined population. Furthermore, the frequencies of people with TT genotype in the LOAD (0.09) and control (0.38) groups were significantly different (p=0.000; OR=0.157; CI: 0.07-0.37). The frequency of heterozygote TC genotype did not show any significant difference between the two investigated groups (p=0.471). These findings suggested that the minor allele T and the genotype TT of the *TOMM40* rs1160985 strongly protect people against the LOAD in the northwestern Iran. These findings are consistent with the reports obtained for the populations of European descent (Roses, 2010), mainland China (Jiao et al., 2015), and the Japanese people (Takei et al., 2009). However, These are inconsistent with the results reported by the studies focused on the Northern-Han Chinese population (Ma et al., 2013). It should be mentioned that the Alzheimer-associated C allele of rs1160985 (*TOMM40*) was reported as the LOAD risk allele by Jiao et al. (Jiao et al., 2015). Unlike the *TOMM40* rs1160985 with protective role against the LOAD, several SNPs of *TOMM40* are served as the LOAD genetic risk factors (Prendecki et al., 2018; Zeitlow et al., 2017).

A high-quality meta-analysis performed on more than 30000 individuals showed that different SNPs in *SORL1* gene are in relationship with the LOAD status (Reitz et al., 2011). We examined the association of an intronic polymorphism of the *SORL1* gene, designated with rs11218304, with the LOAD. Results demonstrated that the frequency of minor allele G is equal to 0.27 in the study population. Comparative studies indicated non-significant differences in the frequency of the allele G in LOAD and control groups (p=0.374). Reversely, the frequency of risk genotype GG was significantly different between the LOAD and control groups (p=0.03; OR=2.053; CI: 0.935-2.053). In fact, although we had 5 patients with GG genotype, we did not observe any individual with the same genotype among the controls. These data, at least to some extent, support similar data from other studies introduced allele G and GG genotype as key genetic risk factors for the appearance of LOAD phenotype in the other populations. For example, Rogaeva et al. reported that rs11218304 (*SORL1*) is significantly associated with LOAD (Rogaeva et al., 2007). Moreover, its association with the poor cognitive efficiency in the LOAD was reported previously (Cruz-Sanabria et al., 2018). However, Ortega-Rojas et al. did not find any significant association between the rs11218304 variant and cognitive decline in the LOAD patients in the Colombians (Ortega-Rojas et al., 2016).

Conclusion

In conclusion, the frequency of rs3810950 (*CHAT*) allele was not significantly different between LOAD and control subjects, while the GG genotype showed a significant association with LOAD in the examined population. Moreover, we observed that the minor allele T of rs1160985 (*TOMM40)* and TT genotype can strongly serve as protective genetic factors against the LOAD. Furthermore, although rs11218304 (*SORL1)* alleles frequencies were not significantly different between the LOAD and control groups, the GG genotype frequency showed a significant difference between the investigated groups. This implies the potential association of GG genotype with the LOAD phenotype in the Azeri population of Northwest Iran.

Acknowledgments

This work was supported by the Cognitive Sciences and Technologies Council of Iran, grant number 5005.

References

Alzheimer's Association. (2019) Alzheimer's disease facts and figures. Alzheimer's & Dementia 15:321-387.

Belmonte Juan Carlos I., Callaway Edward M., Caddick S. J., Churchland P., Feng G., Homanics Gregg E., Lee K.-F., Leopold David A., Miller Cory T., Mitchell Jude F., Mitalipov S., Moutri Alysson R., Movshon J. A., Okano H., Reynolds John H., Ringach D. L., Sejnowski Terrence J., Silva Afonso C., Strick Peter L., Wu J. and Zhang F. (2015) Brains, Genes, and Primates. Neuron 86:617-631.

Bertram L., Lill C. and Tanzi R. (2010) The Genetics of Alzheimer Disease: Back to the Future. Neuron 68:270-281.

Brookmeyer R., Abdalla N., Kawas C. H. and Corrada M. M. (2018) Forecasting the prevalence of preclinical and clinical Alzheimer's disease in the United States. Alzheimer's & Dementia 14:121-129.

Campion D., Charbonnier C. and Nicolas G. (2019) SORL1 genetic variants and Alzheimer disease risk: a literature review and meta-analysis of sequencing data. Acta Neuropathologica 138:173-186.

Cook L., Ho L., Wang L., Terrenoire E., Brayne C., Grimley Evans J., Xuereb J., Cairns N., Turic D., Hollingworth P., Moore P., Jehu L., Archer N., Walter S., Foy C., Edmondson A., Powell J., Lovestone S., Williams J. and Rubinsztein D. (2005) Candidate gene association studies of genes involved in neuronal cholinergic transmission in Alzheimer's disease suggests choline acetyltransferase as a candidate deserving further study. American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics 132B:5-8.

Cruz-Sanabria F., Bonilla-Vargas K., Estrada K., Mancera O., Vega E., Guerrero E., Ortega-Rojas J., Mahecha María F., Romero A., Montañés P., Celeita V., Arboleda H. and Pardo R. (2018) Análisis de desempeños cognitivos y polimorfismos en SORL, PVRL2, CR1, TOMM40, APOE, PICALM, GWAS_14q, CLU y BIN1 en pacientes con trastorno neurocognitivo leve y en sujetos cognitivamente sanos. Neurología.

Dubois B., Feldman H., Jacova C., DeKosky S., Barberger-Gateau P., Cummings J., Delacourte A., Galasko D., Gauthier S., Jicha G., Meguro K., O'Brien J., Pasquier F., Robert P., Rossor M., Salloway S., Stern Y., Visser P. and Scheltens P. (2007) Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria. Lancet neurology 6:734-746.

Felsky D., Szeszko P., Yu L., Honer W. G., De Jager P. L., Schneider J. A., Malhotra A. K., Lencz T., Ikuta T., Pipitone J., Chakravarty M. M., Lobaugh N. J., Mulsant B. H., Pollock B. G., Kennedy J. L., Bennett D. A. and Voineskos A. N. (2014) The SORL1 gene and convergent neural risk for Alzheimer's disease across the human lifespan. Molecular Psychiatry 19:1125-1132.

Fotuhi S. N., Khalaj-Kondori M., Hoseinpour Feizi M. A. and Talebi M. (2019) Long Non-coding RNA BACE1-AS May Serve as an Alzheimer's Disease Blood-Based Biomarker. Journal of Molecular Neuroscience 69:351-359.

Gao L., Zhang Y., Deng J., Yu W. and Yu Y. 2016. Polymorphisms of CHAT but not TFAM or VR22 are Associated with Alzheimer Disease Risk. *In* Medical science monitor : international medical journal of experimental and clinical research. Vol. 22. 1924-1935.

Gatz M., Reynolds C. A., Fratiglioni L., Johansson B., Mortimer J. A., Berg S., Fiske A. and Pedersen N. L. (2006) Role of Genes and Environments for Explaining Alzheimer Disease. Archives of General Psychiatry 63:168-174.

Grupe A., Abraham R., Li Y., Rowland C., Hollingworth P., Morgan A., Jehu L., Segurado R.,

Stone D., Schadt E., Karnoub M., Nowotny P., Tacey K., Catanese J., Sninsky J., Brayne C., Rubinsztein D., Gill M., Lawlor B., Lovestone S., Holmans P., O'Donovan M., Morris J. C., Thal L., Goate A., Owen M. J. and Williams J. (2007) Evidence for novel susceptibility genes for late-onset Alzheimer's disease from a genome-wide association study of putative functional variants. Human Molecular Genetics 16:865-873.

Honea R. A., Vidoni E. D., Swerdlow R. H., Burns J. M. and for the Alzheimer's Disease Neuroimaging I. (2012) Maternal Family History is Associated with Alzheimer's Disease Biomarkers. Journal of Alzheimer's Disease 31:659-668.

Jiao B., Liu X., Zhou L., Wang M. H., Zhou Y., Xiao T., Zhang W., Sun R., Waye M. M. Y., Tang B. and Shen L. (2015) Polygenic Analysis of Late-Onset Alzheimer's Disease from Mainland China. PLOS ONE 10:e0144898.

Lee J., Jo S., Park J., Lee S., Jo I., Kim D., Huh Y., Youn J., Jhoo J., Park K., Park S., Lee D., Woo J. and Kim K. W. (2011) Choline Acetyltransferase 2384G > A Polymorphism and the Risk of Alzheimer Disease. Alzheimer disease and associated disorders 26:81-87.

Liu Y., Chen Q., Liu X., Dou M., Li S., Zhou J., Liu H., Wu Y. and Huang Z. (2016) Genetic Association of CHAT rs3810950 and rs2177369 Polymorphisms with the Risk of Alzheimer's Disease: A Meta-Analysis. BioMed Research International 2016:9418163.

Louwersheimer E., Ramirez A., Cruchaga C., Becker T., Kornhuber J., Peters O., Heilmann S., Wiltfang J., Jessen F., Visser P. J., Scheltens P., Pijnenburg Y. A. L., Teunissen C. E., Barkhof F., van Swieten J. C., Holstege H. and Van der Flier W. M. (2015) The influence of genetic variants in SORL1 gene on the manifestation of Alzheimer's disease. Neurobiology of Aging 36:1605.e1613-1605.e1620.

Lyall D. M., Harris S. E., Bastin M. E., Muñoz Maniega S., Murray C., Lutz M. W., Saunders A. M., Roses A. D., Valdés Hernández M. d. C., Royle N. A., Starr J. M., Porteous D. J., Wardlaw J. M. and Deary I. J. (2014) Are APOE ε genotype and TOMM40 poly-T repeat length associations with cognitive ageing mediated by brain white matter tract integrity? Translational Psychiatry 4:e449-e449.

Ma X.-Y., Yu J.-T., Wang W., Wang H.-F., Liu Q.-Y., Zhang W. and Tan L. (2013) Association of TOMM40 Polymorphisms with Late-Onset Alzheimer's Disease in a Northern Han Chinese Population. NeuroMolecular Medicine 15:279-287.

Miller S. A., Dykes D. D. and Polesky H. F. (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Research 16:1215-1215.

Omoumi A., Fok A., Greenwood T., Sadovnick A. D., Feldman H. H. and Hsiung G.-Y. R. (2014) Evaluation of late-onset Alzheimer disease genetic susceptibility risks in a Canadian population. Neurobiology of Aging 35:936.e935-936.e912.

Ortega-Rojas J., Morales L., Guerrero E., Arboleda-Bustos C. E., Mejia A., Forero D., Lopez L., Pardo R., Arboleda G., Yunis J. and Arboleda H. (2016) Association Analysis of Polymorphisms in TOMM40, CR1, PVRL2, SORL1, PICALM, and 14q32.13 Regions in Colombian Alzheimer Disease Patients. Alzheimer disease and associated disorders 30:305-309.

Petschner P., Gonda X., Baksa D., Eszlari N., Trivaks M., Juhasz G. and Bagdy G. (2018) Genes Linking Mitochondrial Function, Cognitive Impairment and Depression are Associated with Endophenotypes Serving Precision Medicine. Neuroscience 370:207-217.

Prendecki M., Florczak-Wyspianska J., Kowalska M., Ilkowski J., Grzelak T., Bialas K., Wiszniewska M., Kozubski W. and Dorszewska J. (2018) Biothiols and oxidative stress markers and polymorphisms of TOMM40 and APOC1 genes in Alzheimer's disease patients. Oncotarget 9.

Reitz C., Cheng R., Rogaeva E., Lee J. H., Tokuhiro S., Zou F., Bettens K., Sleegers K., Tan E. K., Kimura R., Shibata N., Arai H., Kamboh M. I., Prince J. A., Maier W., Riemenschneider M., Owen M., Harold D., Hollingworth P., Cellini E., Sorbi S., Nacmias B., Takeda M., Pericak-Vance M. A., Haines J. L., Younkin S., Williams J., van Broeckhoven C., Farrer L. A., St George-Hyslop P. H. and Mayeux R. (2011) Meta-analysis of the association between variants in SORL1 and Alzheimer disease. Arch Neurol 68:99-106.

Reynolds C. A., Zavala C., Gatz M., Vie L., Johansson B., Malmberg B., Ingelsson E., Prince J. A. and Pedersen N. L. (2013) Sortilin receptor 1 predicts longitudinal cognitive change. Neurobiology of Aging 34:1710.e1711-1710.e1718.

Rogaeva E., Meng Y., Lee J. H., Gu Y., Kawarai T., Zou F., Katayama T., Baldwin C. T., Cheng R., Hasegawa H., Chen F., Shibata N., Lunetta K. L., Pardossi-Piquard R., Bohm C., Wakutani Y.,

Cupples L. A., Cuenco K. T., Green R. C., Pinessi L., Rainero I., Sorbi S., Bruni A., Duara R., Friedland R. P., Inzelberg R., Hampe W., Bujo H., Song Y.-Q., Andersen O. M., Willnow T. E., Graff-Radford N., Petersen R. C., Dickson D., Der S. D., Fraser P. E., Schmitt-Ulms G., Younkin S., Mayeux R., Farrer L. A. and St George-Hyslop P. (2007) The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. Nature Genetics 39:168-177.

Rosenberg R. N., Lambracht-Washington D., Yu G. and Xia W. (2016) Genomics of Alzheimer Disease: A Review. JAMA Neurology 73:867-874.

Roses A., Sundseth S., Saunders A., Gottschalk W., Burns D. and Lutz M. (2016) Understanding the genetics of APOE and TOMM40 and role of mitochondrial structure and function in clinical pharmacology of Alzheimer's disease. Alzheimer's & Dementia 12:687-694.

Roses A. D. (2010) An Inherited Variable Poly-T Repeat Genotype in TOMM40 in Alzheimer Disease. Archives of Neurology 67:536-541.

Shao W., Peng D. and Wang X. (2017) Genetics of Alzheimer's disease: From pathogenesis to clinical usage. Journal of Clinical Neuroscience 45:1-8.

Takei N., Miyashita A., Tsukie T., Arai H., Asada T., Imagawa M., Shoji M., Higuchi S., Urakami K., Kimura H., Kakita A., Takahashi H., Tsuji S., Kanazawa I., Ihara Y., Odani S. and Kuwano R. (2009) Genetic association study on in and around the APOE in late-onset Alzheimer disease in Japanese. Genomics 93:441-448.

Talebi M., Delpak A., Khalaj-kondori M., Sadigh-Eteghad S., Talebi M., Mehdizadeh E. and Majdi A. (2020) ABCA7 and EphA1 Genes Polymorphisms in Late-Onset Alzheimer's Disease. Journal of Molecular Neuroscience 70:167-173.

Tanzi R. E. (2012) The Genetics of Alzheimer Disease. Cold Spring Harbor Perspectives in Medicine 2.

Thangnipon W., Puangmalai N., Suwanna N., Soi-ampornkul R., Phonchai R., Kotchabhakdi N., Mukda S., Phermthai T., Julavijitphong S., Tuchinda P. and Nobsathian S. (2016) Potential role of N-benzylcinnamide in inducing neuronal differentiation from human amniotic fluid mesenchymal stem cells. Neuroscience Letters 610:6-12.

Willette A. A., Webb J. L., Lutz M. W., Bendlin B. B., Wennberg A. M., Oh J. M., Roses A., Koscik R. L., Hermann B. P., Dowling N. M., Asthana S. and Johnson S. C. (2017) Family history and TOMM40 '523 interactive associations with memory in middle-aged and Alzheimer's disease cohorts. Alzheimer's & Dementia 13:1217-1225.

Yanfang Zhao Y. Z., Lei Zhang, Yanhan Dong, Hongfang Ji, Liang Shen. (2019) The Potential Markers of Circulating microRNAs and long non-coding RNAs in Alzheimer's Disease. Aging and disease 10:1293-1301.

Yavarpour-Bali H., Ghasemi-Kasman M. and Shojaei A. (2020) Direct reprogramming of terminally differentiated cells into neurons: A novel and promising strategy for Alzheimer's disease treatment. Progress in Neuro-Psychopharmacology and Biological Psychiatry 98:109820.

Yu C.-E., Seltman H., Peskind E. R., Galloway N., Zhou P. X., Rosenthal E., Wijsman E. M., Tsuang D. W., Devlin B. and Schellenberg G. D. (2007) Comprehensive analysis of APOE and selected proximate markers for late-onset Alzheimer's disease: Patterns of linkage disequilibrium and disease/marker association. Genomics 89:655-665.

Yu X., Li Y., Wen H., Zhang Y. and Tian X. (2015) Intensity-dependent effects of repetitive anodal transcranial direct current stimulation on learning and memory in a rat model of Alzheimer's disease. Neurobiology of Learning and Memory 123:168-178.

Yuan H., Xia Q., Ling K., Wang X., Wang X. and Du X. (2016) Association of Choline Acetyltransferase Gene Polymorphisms (SNPs rs868750G/A, rs1880676G/A, rs2177369G/A and rs3810950G/A) with Alzheimer's Disease Risk: A Meta-Analysis. PLOS ONE 11:e0159022.

Zeitlow K., Charlambous L., Ng I., Gagrani S., Mihovilovic M., Luo S., Rock D. L., Saunders A., Roses A. D. and Gottschalk W. K. (2017) The biological foundation of the genetic association of TOMM40 with late-onset Alzheimer's disease. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 1863:2973-2986.

Scientific Reviewers

Bahar Shahnavaz, Ph.D., (Assistant Professor of Microbiology), Ferdowsi University of Mashhad, Mashhad, Iran

Sara Soltanian, Ph.D., (Assistant Professor of Cell and Molecular Biology), Shahid Bahonar University of Kerman, Kerman, Iran

Sara Yousefi Taemeh, Ph.D. scholar, (Ferdowsi University of Mashhad, Mashhad, Iran)

MANUSCRIPT PREPARATION

Manuscripts should be prepared in accordance with the uniform requirements for Manuscript's Submission to **"Journal of Cell and Molecular Research".**

Language: Papers should be in English (either British or American spelling). The past tense should be used throughout the results description, and the present tense in referring to previously established and generally accepted results. Authors who are unsure of correct English usage should have their manuscript checked by somebody who is proficient in the language; manuscripts that are deficient in this respect may be returned to the author for revision before scientific review.

Typing: Manuscripts must be typewritten in a font size of at least 12 points, double-spaced (including References, Tables and Figure legends) with wide margins (2.5 cm from all sides) on one side of the paper. The beginning of each new paragraph must be clearly indicated by indentation. All pages should be numbered consecutively at the bottom starting with the title page.

Length: The length of research articles should be restricted to ten printed pages. Short communication should not exceed five pages of manuscript, including references, figures and tables. Letters should be 400-500 words having 7-10 references, one figure or table if necessary. Commentaries and news should also be 800-1000 words having 7-10 references and one figure or table if necessary.

Types of Manuscript: JCMR is accepting original research paper, short communication reports, invited reviews, letters to editor, biographies of scientific reviewers, commentaries and news.

Statement of Human and Animal Rights: Author's should declare regulatory statement regarding the experiments using animals, human cells/tissues that all in vivo experiments have been performed according to the guidelines (explained by WHO, international animal rights federations or your respective institute) to use animals in their research work.

Conflict of Interest Statement: Authors or corresponding author should declare statement of conflict of interest at the last of manuscript.

Manuscript Evaluation Time: All submitted manuscripts will be evaluated and reviewed according to following evaluation schedule.
Pre-Editorial Evaluation: All submitted manuscripts, right after their submission to JCMR will be evaluation by Editors for being according to the journal scope and format. This evaluation can take 2-7 days of submission.
Reviewer's Evaluation: Selected manuscripts after pre-editorial evaluation will be sent to minimum two blind reviewers assigned by Editor-in-Chief. This process may take 21-27 days.
Post Editorial Evaluation: After receiving reviewer's comments, editors evaluate the manuscripts considering the comments and decide their first decision. This process takes 3-5 days and then authors are informed regarding the editorial decision.

GENERAL ARRANGEMENT OF PAPERS

Title: In the first page, papers should be headed by a concise and informative title. The title should be followed by the authors' full first names, middle initials and last names and by names and addresses of laboratories where the work was carried out. Identify the affiliations of all authors and their institutions, departments or organization by use of Arabic numbers (1, 2, 3, etc.).

Footnotes: The name and full postal address, telephone, fax and E-mail number of corresponding author should be provided in a footnote.

Abbreviations: The Journal publishes a standard abbreviation list at the front of every issue. These standard abbreviations do not need to be spelled out within paper. However, non-standard and undefined abbreviations used five or more times should be listed in the footnote. Abbreviations should be defined where first mentioned in the text. Do not use abbreviations in the title or in the Abstract. However, they can be used in Figures and Tables with explanation in the Figure legend or in a footnote to the Table.

Abstract: In second page, abstract should follow the title (no authors' name) in structured format of not more than 250 words and must be able to stand independently and should state the Background, Methods, Results and Conclusion. Write the abstract in third person. References should not be cited and abbreviations should be avoided.

Keywords: A list of three to five keywords for indexing should be included at bottom of the abstract. Introduction should contain a description of the problem under investigation and a brief survey of the existing literature on the subject.

Materials and Methods: Sufficient details must be provided to allow the work to be repeated. Correct chemical names should be given and strains of organisms should be specified. Suppliers of materials need only be mentioned if this may affect the results. Use System International (SI) units and symbols.

Results: This section should describe concisely the rationale of the investigation and its outcomes. Data should not be repeated in both a Table and a Figure. Tables and Figures should be selected to illustrate specific points. Do not tabulate or illustrate points that can be adequately and concisely described in the text.

Discussion: This should not simply recapitulate the Results. It should relate results to previous work and interpret them. Combined Results and Discussion sections are encouraged when appropriate.

Acknowledgments: This optional part should include a statement thanking those who assisted substantially with work relevant to the study. Grant support should be included in this section.

References: References should be numbered and written in alphabetical order. Only published, "in press" papers, and books may be cited in the reference list (see the examples below). References to work "in press" must be accompanied by a copy of acceptance letter from the journal. References should not be given to personal communications, unpublished data, manuscripts in preparation, letters, company publications, patents pending, and URLs for websites. Abstracts of papers presented at meetings are not permissible. These references should appear as parenthetical expressions in the text, e.g. (unpublished data). Few example of referencing patterns are given as follows:

Bongso A., Lee E. H. and Brenner S. (2005) Stem cells from bench to bed side. World Scientific Publishing Co. Singapore, 38-55 pp.

Irfan-Maqsood M. (2013) Stem Cells of Epidermis: A Critical Introduction. Journal of Cell and Molecular Research 5(1): 1-2.

Note: All the reference should be in EndNote format (JCMR EndNote Style is available on JCMR's web site, Author's Guideline)

Tables and Figures: Tables and Figures should be numbered (1, 2, 3, etc.) as they appear in the text. Figures

should preferably be the size intended for publication. Tables and Figures should be carefully marked. Legends should be typed single-spaced separately from the figures. Photographs must be originals of high quality. Photocopies are not acceptable. Those wishing to submit color photographs should contact the Editor regarding charges.

Note: All manuscripts submitted to JCMR are tracked by using "Plagiarism Tracker X" or "small tools" for possible plagiarism before acceptance to JCMR.

Table of Contents

Journal of Cell and Molecular Research

Volume 12, Number 1, Summer, 2020