

In the Name of Allah,
The All-beneficent, the All-merciful

UNIVERSITÀ DEGLI STUDI DI TORINO



Department of Life Sciences and Systems Biology

Ph.D. Program in Complex Systems for Life Sciences

Cycle: XXXIV

Transcriptome Profiling of Human Embryonic Stem Cells Differentiation to Skeletal Muscle
Cells and Involvement of Cancer Genes, and Transcription and Epigenetic Factors

Thesis' author: Hassan Dastsooz

Tutor: Prof. Salvatore Oliviero

Academic field: BIO/11 BIOLOGIA MOLECOLARE
Academic years of enrolment: 2018-2021

Contents

ACKNOWLEDGEMENT.....	5
Part one	6
Abstract.....	7
Chapter1.....	9
INTRODUCTION.....	9
1.1 Human embryonic stem cells (hESCs) and its applications.....	9
1.2 hESCs differentiation to muscle progenitors and myocytes.....	11
1.3 Myokines involvement in muscle cells differentiation.....	13
1.4 Role of ubiquitin proteins, mainly <i>UBE2C</i> , in muscle differentiation	15
1.5 Genes affecting OS of patients with breast cancer and their involvement in muscle cells differentiation.....	18
Chapter 2.....	19
MATERIAL AND METHODS.....	19
2.1 H9 hESCs differentiation to skeletal myogenic progenitors and skeletal muscle cells.....	19
2.1.1 H9 hESCs culture using feeder-free culture system.....	19
2.1.2 hESCs morphology.....	19
2.1.3 EDTA-based passaging of H9 hESCs.....	21
2.2 Differentiation into the myogenic lineage.....	22
2.2.1 Culture of H9 hESCs.....	22
2.2.2 Differentiation procedure.....	22
2.2.3 Differentiation (Day 0–46)	23
2.2.4 IF staining for 8-well chamber slide.....	25
2.2.5 QRT-PCR analysis.....	26
2.2.5.1 RNA extraction by QiAzol reagent.....	26
2.2.5.2 QRT-PCR.....	27
2.2.6 RNA sequencing library preparation for transcriptome sequencing.....	28
2.2.7 RNA Sequencing Analysis.....	28
2.3 Expression of TCGA-breast adenocarcinoma (BRCA) genes affecting OS across muscle cell differentiation	29
2.4 Investigation of ubiquitin proteins, mainly <i>UBE2C</i> , in muscle differentiation stages.....	30
Chapter 3.....	33

RESULTS.....	33
3.1 Confirmation of H9 differentiation to skeletal myogenic progenitors and skeletal muscle cells using stage markers.....	33
3.2 Transcriptome profiling of muscle differentiation stages.....	39
3.2.1 RNA sequencing quality control and Principal Component Analysis (PCA).....	39
3.2.2 Confirmation of stage specific markers.....	42
3.2.3 Differential gene expression and time-course gene expression profiling.....	44
3.2.4 Cluster profiling based on GO: Biological Processes and KEGG pathways.....	45
3.3 Involvement of genes with/out transcription factors (TFs) and epigenetic factors (Epifactors) functioning as proto-oncogenes and tumor suppressors across muscle differentiation stages.....	47
3.3.1 Cancer drivers with proto-oncogenic and tumor suppressive roles from cluster 8.....	48
3.3.2 Cancer drivers with TF role as proto-oncogenes from clusters 4 and 8.....	50
3.3.3 Involvement of cancer drivers with only TF role as proto-oncogenes across muscle differentiation stages.....	51
3.3.4 Cancer drivers with Epifactor role with proto-oncogenic function across muscle differentiation stages.....	52
3.3.5 Cancer drivers with both TF and Epifactor roles with both proto-oncogenic and tumor suppressive functions across muscle differentiation stages among clusters 4 and 7.....	53
3.3.6 Cancer driver genes with both tumor suppressive and proto-oncogenic roles which function as TF and/or Epifactor among clusters 4,7, and 8.....	54
3.3.7 Cancer driver genes with only TF role which function as tumor suppressors.....	55
3.3.8 Cancer drivers with Epifactor role which function as tumor suppressors.....	55
3.3.9 Cancer drivers with roles of both TF and Epifactor which function as tumor suppressors from cluster 4.....	56
3.3.10 Tumor suppressor genes with TF role.....	56
3.3.11 Tumor suppressor genes with Epifactor role.....	56
3.3.12 Tumor suppressor genes with both Epifactor and TF roles.....	58
3.3.13 Genes with roles of both TF and Epifactor which function as both proto-oncogenes and tumor suppressors across all clusters.....	59
3.3.14 Proto-oncogenes without TF and Epifactor roles.....	60
3.3.15 Tumor suppressor genes without TF and Epifactor roles.....	61
3.3.16 Tumor suppressor genes and proto-oncogenes without TF and Epifactor roles among all clusters.....	62

3.3.17 The involvement of genes with role as both TF and Epifactor in muscle differentiation stages.....	63
3.3.18 The involvement of genes with role as only Epifactor in muscle differentiation stages....	67
3.4 Involvement of myokines in muscle differentiation.....	73
3.5 Involvement of breast cancer genes in muscle differentiation.....	75
3.6 Involvement of ubiquitin proteins in muscle differentiation and cancer.....	80
3.7 <i>UBE2C</i> involvement in tumors and muscle differentiation.....	84
Chapter 4.....	86
DISCUSSION.....	86
REFERENCES.....	94
Part two	99
PUBLICATIONS.....	99
Abstract.....	100
Article 1: Expanding the molecular and clinical phenotypes of FUT8-CDG	101
Article 2: A Comprehensive Bioinformatics Analysis of <i>UBE2C</i> in Cancers	112
Article 3: Truncating Mutations in UBAP1 Cause Hereditary Spastic Paraplegia	127
Article 4: LncRNA-miRNA-mRNA Networks of Gastrointestinal Cancers Representing Common and Specific LncRNAs and mRNAs.....	133
Article 5: Autophagy: The Potential Link between SARS-CoV-2 and Cancer.....	142
Article 6: Clinical and molecular characterization of a patient with mitochondrial Neurogastrointestinal Encephalomyopathy.....	151
Article 7: A Novel TTC19 Mutation in a Patient With Neurological, Psychological, and Gastrointestinal Impairment.....	157
Article 8: Clinical and molecular characterization of three patients with Hepatocerebral form of mitochondrial DNA depletion syndrome: a case series.....	162
Article 9: A novel frame-shift deletion in FANCF gene causing autosomal recessive Fanconi anemia: a case report.....	171
Article 10: A novel mutation in SEPN1 causing rigid spine muscular dystrophy 1: a Case report...177	
Article 11: Clinical Values of two Novel Estrogen Receptor Signaling Targeted lncRNAs in Invasive ductal breast carcinoma.....	182
SUPPLEMENTAL INFORMATION.....	191

ACKNOWLEDGEMENT

I especially would first like to thank my tutor, Professor Salvatore Oliviero, who has provided me the opportunity to pursue my Ph.D. and work at this lab and for his patience, guidance, and support. I am extremely grateful that he took me on as a student and continued to have faith in me over the years. His invaluable expertise formulates my research hypothesis and methodology, and his insightful feedbacks pushed me to sharpen my thinking and brought my work to a higher level.

I acknowledge the generous financial support from the University of Torino.

Very special thanks to my group colleagues for their excellent collaboration. I would like to thank Valentina Proserpio and Francesca Anselmi for their kind supports during my projects and all of the solutions to the experiment issues to further my research. I want to thank Andrea Lauria and Chiara Cicconetti for their helpful data analysis. Besides, I want to thank other colleagues, Fatemeh, Giulia, Hua, Corinne, Isabelle, Carlo, Mara, Stefania, Matteo, Annalaura, Mirko, Ivan, Daniela, Lisa, Caterina, Danny, Edoardo, Daniele, Claudia, and all people in MBC and IIGM for creating a pleasant lab atmosphere.

In addition, I would like to thank my family for their wise counsel and sympathetic ear. You are always there for me. Finally, I have to thank my wife and love of my life, Farideh, for keeping things going and for always showing how proud you are of me and for the sacrifices you have made in order for me to pursue my degree. I simply couldn't have done this without you.

Thanks also to those who are reading my thesis, and I hope you enjoy our work.

وَمَا لِي لَا أَعْبُدُ الَّذِي فَطَرَنِي وَإِلَيْهِ تُرْجَعُونَ

آخر چرا کسی را نپرستم که مرا آفریده است و [همه] شما به سوی او بازگشت می یابید (سوره یاسین، آیه ۲۲).

Why should I not worship Him who has originated me, and to whom you shall be brought back?

Part One

Transcriptome Profiling of Human Embryonic Stem Cells Differentiation to Skeletal Muscle Cells and Involvement of Cancer Genes, and Transcription and Epigenetic Factors

ABSTRACT

Human embryonic stem cells (hESCs) have remarkable capability to indefinitely self-renew in the undifferentiated state and the capacity to differentiate into all three embryonic germ layers. These cells have an enormous potential as a source for cell-based therapies and disease model systems. Recently, there has been of great interest to use myogenic progenitors differentiated from hESCs for regenerative medicine. Moreover, recent studies have provided the involvement of skeletal muscle cells in cancer, mainly breast tumor. Therefore, great interests have been made for identification of pathways related to the hESCs differentiation to muscle cells with different approaches. In our study, the main aim was to investigate the gene expression profiling across muscle differentiation derived from hESCs. We used a previously described protocol with some minor modifications to achieve and expand *PAX7*⁺ expressing skeletal muscle precursors and also myocytes from H9 hESCs, helping investigate biological processes and pathways involved in muscle differentiation.

We successfully differentiated H9 hESCs to muscle cell and its progenitors (skeletal muscle progenitors, SMPs) with confirmation of high expression of stage markers for each stage. Our RNA sequencing data revealed 8641 differential expressed genes in nine distinct clusters over the muscle differentiation stages. Biological process and KEGG pathway profiling revealed different terms for each muscle differentiation stage which were in concordance with the stages. Moreover, our data identified the altered expression of a number of genes with different functions as transcription factors (TFs) and epigenetic factors (Epifactors), proto-oncogenes, and tumor suppressors among others.

Among genes with only Epifactor role with proto-oncogenic function, almost all of them were downregulated across almost all stages, representing their involvement in stemness status of hESCs and cancers. Regarding tumor suppressor genes with TF role, interestingly almost all of them revealed higher expression at almost all muscle differentiation stages, which is consistent with their tumor suppressive function. In relation to the tumor suppressor genes with only Epifactor role, we identified that most of these genes were upregulated as expected for such genes.

Regarding DNA methyltransferases, our data revealed lower expression of *DNMT3B* at all stages compared to stemness stage, representing the main role of *DNMT3B* in stemness status. However, *DNMT3A* only had lower expression at mesoderm stage and *DNMT1* only at late stages.

In another category for genes with both tumor suppressive and proto-oncogenic roles but without TF and Epifactor functions, our results showed expression alterations of various important genes, in which most of them had upregulation during muscle differentiation stages, involving several main cancer genes such as *EPHA3*, *CDH11*, *CDKN2C*, *DCC*, *CDKN2A*, among others.

Moreover, we found that some *FOX* genes such as *FOXA1*, *FOXP1*, *FOXP2*, and *FOXP4* with role in transcription regulation showed upregulation from somite stage toward the late stage.

Also, our data revealed involvement of main myokines in muscle differentiation, with their upregulation across the muscle stages, confirming their expected role in reduction of stemness status of cancer and stem cells.

Since altered expression of genes involved in muscle cell homeostasis has been found in pectoralis muscle samples from breast tumors, we investigated and found the involvement of a number of The Cancer Genome Atlas (TCGA)-breast cancer (BRCA) genes across muscle differentiation. We also could show involvement of some uncharacterized and possible novel tumor suppressors from TCGA BRCA in muscle differentiation, mainly for *BOC*, *PLEKHA4*, *BEGAIN*, and *N4BP2L1* with their higher significant expression.

In relation to the involvement of ubiquitin proteins in muscle differentiation, we found the significant altered expression of a number of genes, with higher expression at stemness stage compared to late differentiated stages. Moreover, for the first time, we found a highly constant expression of *UBAPI*, impaired in the hereditary spastic paraplegia, during all muscle stages, indicating its possible important role during all muscle differentiation stages. In addition, we found the involvement of *UBE2C* and its co-expressed genes from TCGA cancers in muscle differentiation stages. While *UBE2C* had significant overexpression in TCGA cancers, it showed significant lower expression at late stages of muscle differentiation compared to stemness stage, indicating its function as proto-oncogene to maintain stemness status of the cancer and ES cells.

In summary, our hESCs differentiation toward muscle cells could help uncover a huge number of known and unknown genes as TF and/or Epifactors which also have proto-oncogenic and/or tumor suppressive roles.

Chapter 1

INTRODUCTION

1.1 Human embryonic stem cells (hESCs) and its applications

Human pluripotent cells (hPSCs) include embryonic stem cells (hESC), and human induced pluripotent cells (hiPSCs). hESC is originated from the inner cell mass (ICM) of the blastocyst while hiPSCs is generated from a direct reprogramming of human somatic cells to have function like embryonic stem cells by forced and transgenic expression of specific transcription and reprogramming factors [1, 2].

Embryonic stem (ES) cells have several key characteristics as follow: A) They are originated from a population of pluripotent cells. B) ES cells have the remarkable capability to indefinitely self-renew in the undifferentiated state. C) They maintain both normal karyotype and markers of undifferentiated cells such as Oct-4, NANOG, SOX2, and tumor rejection antigens (Tra-1-81 and Tra-1-60) through prolonged culture. D) ES cells have the capacity to differentiate into all three embryonic germ layers in vitro or to form teratomas in vivo. (Figure 1). E) They show significant positive signal for cell membrane surface markers SSEA 3 and SSEA 4 (stage-specific-embryonic antigens) while negative for SSEA 1 [2, 3].

hESCs are a robust cell source and flexible model system to study a number of tissues. Using an appropriate stimulant, hESCs are able to differentiate into given cell types and tissues. Beside their applications for studding human biological pathways, disease mechanisms, and drug screenings, these PSCs as in vitro model have a remarkable controllable proliferation capacity to yield a sufficient quantity of transplantable material for cell-based therapeutics [1, 4].

The aim of our study is to differentiate H9 hESCs into the skeletal muscle progenitors and myocytes and also investigate the transcriptome profiling of the differentiation's stages, helping uncover genes, pathways and biological processes involved this process. Using the RNA expression profile data, in

our feature studies, we will consider functional studies for some genes to reveal their role in skeletal muscle differentiation and cancer.

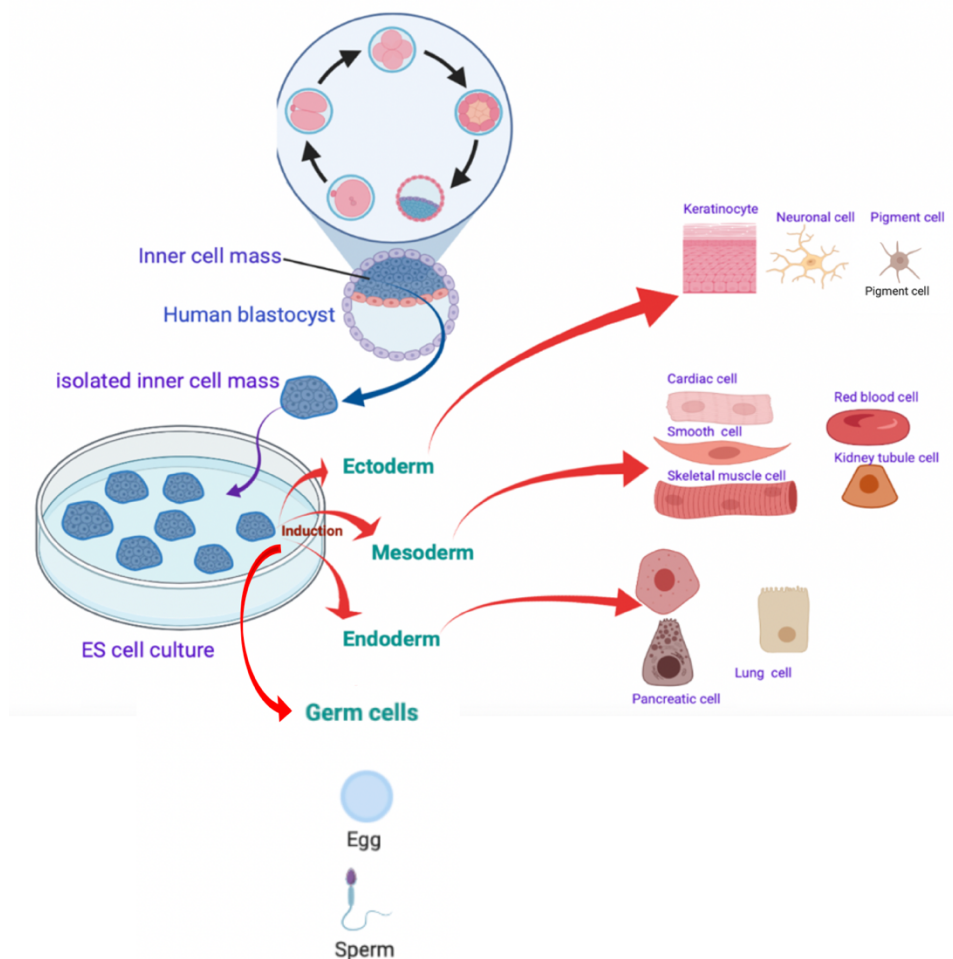


Figure 1. Deriving hESC line and its differentiating into different human cell types. An hESC line is generated from the inner cell mass of a blastocyst and maintained in ES medium. The provided cell line has potential ability to be differentiated into several cell types originated from the different embryonic lineages. The differentiation can be initiated by different approaches such as embryoid body formation and induction with some components and factors. The figure was generated using [Biorender.com](https://www.biorender.com), accessed on 18 and 19 March 2022.

1.2 hESCs differentiation to muscle progenitors and myocytes

hESCs differentiation to skeletal muscle cells needs different stages to reach terminal differentiation as follow: 1-Paraxial mesoderm [expressing Brachyury (*T*), Mesogenin 1 (*MSGN1*), and *TBX6*], 2-Somite stage [expressing *PAX3*, *MEOX1*, and *PAX7*], 3-Skeletal muscle progenitors (SMPs) [expressing the MRFs (*MYF5*, *MYOD1*, and *MYOG*), and *PAX7*], and 4-Terminal differentiation [expression of the MRFs and muscle structural protein (Myosin heavy chain proteins such as MYH3 and MF20)] [4] (Figures 2). The terminal muscle differentiation occurs when myogenic progenitors expressing *PAX3* and/or *PAX7* start to produce muscle determination genes, *MYF5* or *MYOD*, to be as committed myoblasts [5-7]. Myogenin (*MYOG*) expression emerge gradually in these cells and they begin to establish myosin heavy chain (MHC)-expressing nascent myotubes (*MHC+*) [4, 5](Figure 2). Induction of myotube fusion for multinucleated myotube formation can be initiated using some factors such as Insulin-like growth factor-I (IGF-I); and inhibitors of TGF- β 1 and myostatin. Then, organized sarcomeres (the smallest functional contractile unit of a skeletal muscle fiber) within the myotubes are established by presence and arrangement of actin, myosin, and elastic myofilaments [8-11].

In our study, we use a protocol described by Shelton, M et al. [4] (Figure 2) and will provide SMPs, committed myocyte and nascent myotube with the advantage of cell-based therapy to uncover the profile expression of genes, mainly cancer genes, transcription factors and epigenetic factors during muscle differentiation stages. Skeletal muscle differentiation from hESCs involves induction of paraxial mesoderm by GSK3 (CHIR99021) Inhibitor which activates Wnt signaling and high expression of the key transcription factors at this stage confirms the presence of earliest paraxial mesodermal cells with potential ability to differentiate toward skeletal myogenic lineage [4]. Then, these cells are highly proliferated to give rise to somite stage. Next, using fibroblast growth factor 2 (FGF2) supplementation, cells are committed to the myogenic lineage with favored SMPs expansion, with higher expression levels of the MRFs than *PAX7* which confirm proceeding with the early dermomyotome-like stage [4]. The first differentiated myocytes are presented between D30-35 for

H9 hESCs which is dependent on the stem cell line. In the passaging protocol [4], we use a given media supplemented with FGF2 from SMPs stage (D20) upward to achieve terminal differentiation and maturation of the myogenic cultures along with the PAX7-expressing population, helping expand cells after each passaging. This approach is beneficial for cell-based therapy since we are able to provide sufficient numbers of skeletal muscle precursors and myocytes for further studies and investigations.

In our study, we use this well-defined protocol for 46 days to generate and expand skeletal muscle precursors and myocytes from H9 hESCs. One of the advantages of this protocol is the presence of SMPs since in cell therapy strategies terminally differentiated myoblasts or myocytes does not have ability to generate and proliferate the new cells and they have limited repair capacity following transplantation. However, provision of the dividing SMPs help differentiate them toward new myocytes and maintain the PAX7⁺SMPs pool interspersed among skeletal myocytes [4]. Therefore, in current study we will differentiate hESC cells to muscle cells and, using RNA sequencing, we investigate the expression of cancer genes; transcription and epigenetic factors; and their pathways during its different stages.

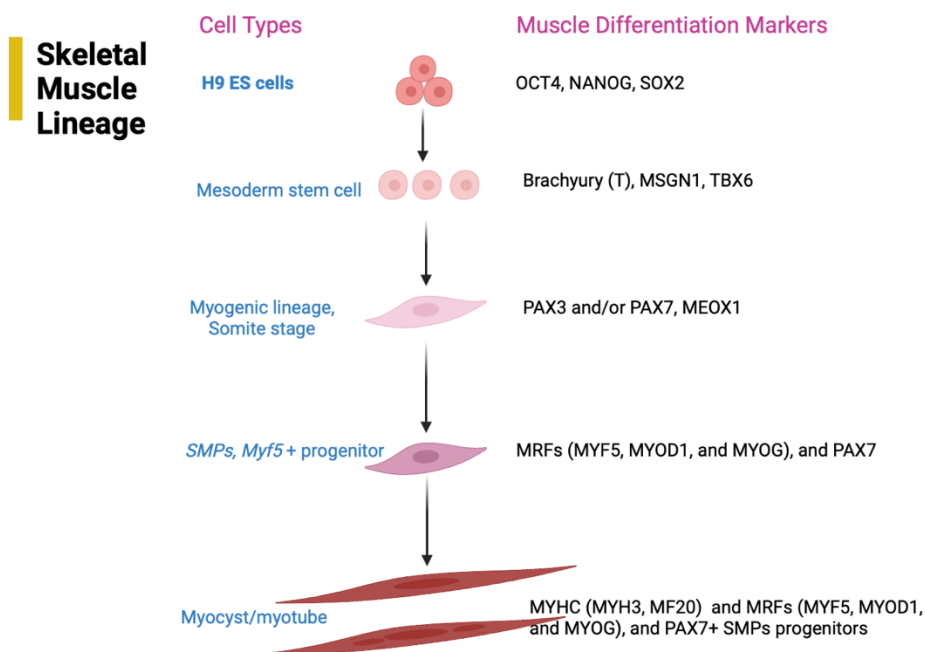


Figure 2. The overview of skeletal muscle differentiation from hESCs. The figure was generated using [Biorender.com](https://biorender.com), accessed on 19 March 2022.

1.3 Myokines involvement in muscle cells differentiation

During the past decades, an increasing advances have been made in stem cell biology, helping focus mainly on its applications and use in treatment and generation of disease models, including neuromuscular disorders [12]. These diseases are caused by muscle pathology or dysfunction of the peripheral nerves. However, so far, a promising treatment approach for most of them has not been developed. Therefore, replacing damaged skeletal muscle cells with normal myocytes or their myogenic progenitors generated from ES cells represents an effective therapeutic strategy for these disorders [8, 13]. Moreover, based on recent studies representing involvement of skeletal muscle cells in cancer (skeletal muscle as an endocrine organ can provide anti-inflammatory and anti-cancer factors[14-16]) and the stemness characteristics of cancer (cancer stemness is a major driver of cancer progression and its survival [17]), an increasing interest has been made in identification of pathways and networks related to differentiation of hESCs to muscle cells. This differentiation process can help identify cancer related genes at its stages. Some studies have shown that regular physical activities could result in the reduced risk of developing a variety of cancers, including breast cancer, and decrease the risk of cancer recurrence and also better overall survival (OS) in patients with cancers [14-16].

As mentioned above, skeletal muscle is an endocrine organ with anti-inflammatory and anti-cancer effects and some studies identified that skeletal muscle cells produce, express, and release some humoral factors (called myokines) [16, 18]. However, biological roles of the majority of these myokines have not been identified. Studies have shown that specific myokines could produce and release upon physical activities and could provide a healthier status involving anti-inflammatory factors and various processes in muscle tissues in patient with cancers [18-21]. Furthermore, studies on cancer cell lines have shown that serum from cancer patients who had exercises and physical activities could reduce human tumor cell viability. Based on this finding, it was proposed that these activities could have impacts on specific gene expression regulating the proliferation and apoptosis in human cancer cell lines, representing the anti-tumor effects of myokines [22]. The role of some of

these myokines are described in Table 1 [23]. Interleukin 6 (IL6) is one of the major myokines which is immediately produced and released from contracting skeletal muscle and resulted in the anti-inflammation effects [20, 24, 25]. It has been shown that IL-6 released after physical activities could have an anti-inflammatory function through an inhibitory effect [26-28]. Moreover, myokines could have inhibitory role in inhibiting mammary cancer cell proliferation and viability by apoptosis induction or tumor-fighting immune system stimulation [15]. In addition, it has been reported that myokines such as *IL10*, *CXCL1* and *CCL4* could upregulate caspase 3 and 7 expression in cancer cells, leading to the increased cancer cell death [21].

Based on these findings and the possible role of myokines in cancers, mainly muscle-related cancers, in one part of study, we look for involvement of some myokines in muscle differentiation stages, identifying uncharacterized biological and oncogenic pathways in muscle differentiation stages.

Table 1. Functions of myokines upon physical activities.

	Increase	Decrease
IL-6	Muscle atrophy and oxidation of fatty acid	Inflammation
IL-15	Metabolism of fat, differentiation of myoblast, and muscle mass/atrophy of muscle	
BDNF	Regeneration of muscle and oxidation of fatty acid	
SPARC	Repair of muscle	
FGF21	Muscle mass and biogenesis of mitochondrial	
DECORIN	Myogenesis	Muscle atrophy
MYONECTIN	Biogenesis of mitochondria	Autophagy
MYOSTATIN	Muscle atrophy	Muscle mass
IRISIN	Muscle mass, oxidation of fatty acid, and muscle hypertrophy	

1.4 Role of ubiquitin proteins, mainly *UBE2C*, in muscle differentiation

Skeletal muscle which is contributed nearly forty percent of human body weight (Figure 3A) is one of the most dynamic tissues of our body consisted of approximately 75% of all proteins in the body. Muscle mass is depended on the balance between protein synthesis and degradation and these processes are affected by physiological and pathological conditions [29, 30].

Physiological homeostasis (proteostasis) of skeletal muscle is relied on various signaling pathways controlled by extracellular and intracellular signals. Protein homeostasis involves the essential balance between cellular protein synthesis, maintenance and degradation. The two major machineries responsible for intracellular protein degradation include the ubiquitin proteasome system (UPS) and the autophagy-lysosomal pathway (ALP). ALP involves lysosomal proteolytic enzymes for degradation of cytoplasmic proteins and organelles involved by autophagosomes while the UPS is responsible for the targeted degradation of both cytosolic and nuclear proteins with the involvement of the ubiquitin/proteasome proteins. Muscle protein degradation in physiological and pathological conditions is occurred by coordinately functions of the UPS, the ALP, and proteases such as calpains and caspases [29, 31, 32].

Pectoralis muscle attaches the front of chest walls to the upper extremities (Figure 3B). On each side of the breastbone, two such muscles are present which include pectoralis major and pectoralis minor [33]. In a study conducted by Naziya Samreen et al., they showed that pectoralis muscle enhancement in terms of mass tissue was observed in most of their samples with breast tumors [34]. Moreover, two studies on pectoralis muscle from breast cancer patients with early stage non-metastatic condition revealed an increased expression of genes related to ubiquitin-mediated proteolysis (increased the protein degradation) and a decreased expression of genes related to ribosomes (decreased protein synthesis), indicating a change in protein homeostasis and predisposing them to muscle fatigue (impaired muscle functions but not necessarily loss of the muscle mass) [35].

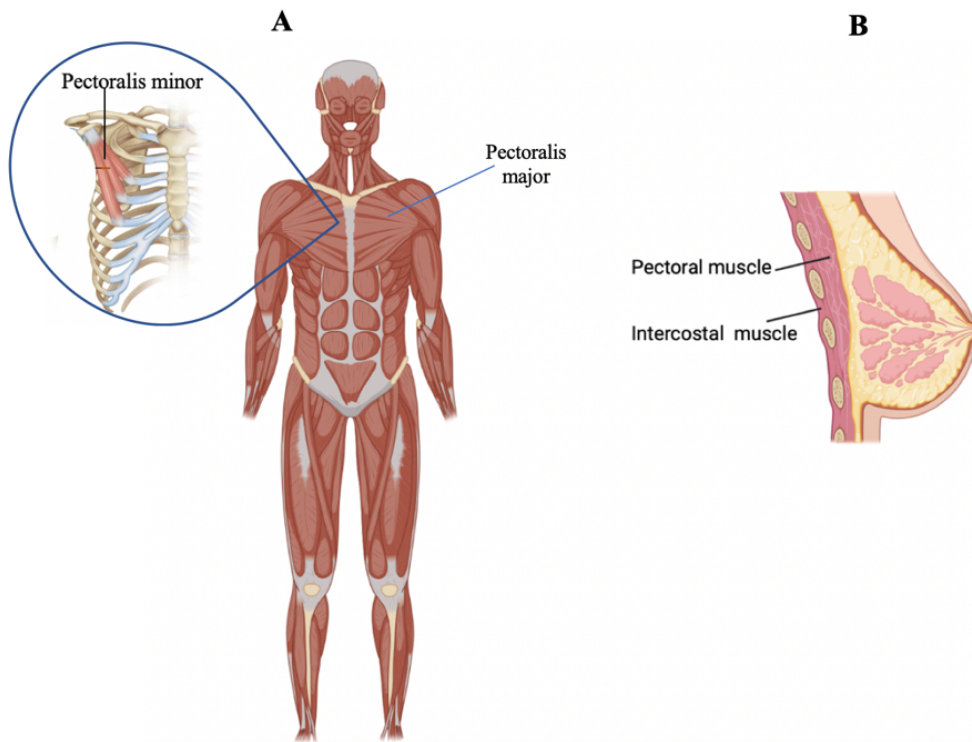


Figure 3. **A.** Overview of the presence of skeletal muscle in human body. **B.** Pectoralis muscle in breast tissue. The figure was generated using Biorender.com, accessed on 18 March 2022 and also one part of 3A was extracted from following book: Warth R.J., Millett P.J. (2015) Strength Testing. In: Physical Examination of the Shoulder. Springer, New York, NY. https://doi.org/10.1007/978-1-4939-2593-3_3.

Moreover, as mentioned above pectoralis muscle of breast cancer patients showed an elevated expression of genes involved ubiquitin-mediated proteolysis and; therefore, based on these evidences, in another part of our study, we look for the expression of ubiquitin proteins, mainly *UBE2C* and its co-expressed genes extracted from The Cancer Genome Atlas (TCGA), across muscle differentiation. *UBE2C* has been reported to be overexpressed in all cancers including breast cancer, reported by Dastsooz et al. [36] and other researchers in different cancers[37-43].

Ubiquitin, a highly conserved small eukaryotic protein, has a vital role in controlling substrate degradation. Ubiquitination is one of the evolutionarily conserved protein post-translational modification and polyubiquitination chain is occurred when one or multiple defined lysine (K) residues of ubiquitin or its N-terminal methionine residue (M1) is initially ubiquitinated [36, 44-46]. This process involve a multi-step process mediated by three different enzymes, including E1 ubiquitin-activating enzymes, E2 ubiquitin-conjugating enzymes, and E3 ubiquitin-ligating enzymes. This system results in the proteasomal removal of its substrate or influence the substrate activity, localization, and interactions with other proteins [36, 45, 47, 48].

In the ubiquitin-proteasome system, anaphase promoting complex/cyclosome (APC/C) is mainly relied on the ubiquitin conjugating enzyme, E2C (*UBE2C*), for priming the ubiquitin chain formation on APC/C substrates, and also *UBE2S*, for extending and branching the ubiquitin. This system is essential to regulate protein homeostasis and control fundamental cellular processes; therefore, it is expected its impaired function leads to human diseases, including cancers [49, 50].

In our previously reported study, we provided evidences that *UBE2C* could act as a proto-oncogene and its overexpression correlated with worsen OS, suggesting its role in progression and invasion of various tumors. However, across muscle differentiation, we observe downregulation of *UBE2C* and its co-expressed genes in our RNA sequencing data during late stages of muscle differentiation, representing their main role in stemness status of hESCs and cancers.

1.5 Genes affecting OS of patients with breast cancer and their involvement in muscle cells differentiation

Using data of TCGA BRCA, we found many genes with altered expression affecting OS in breast cancer. Therefore, based on the issue that stem cells (SCs) and cancer stem cells (CSCs) share common regulatory signaling pathways, we looked for the involvement of these genes mainly their uncharacterized and possible novel tumor suppressors in muscle differentiation, helping identify their corresponding pathways.

Chapter 2

MATERIA and METHODS

2.1 H9 hESCs differentiation to skeletal myogenic progenitors and skeletal muscle cells

2.1.1 H9 hESCs culture using feeder-free culture system

In our study, we used H9 hESCs (a normal female karyotype 46, XX). The cells were grown and maintained in TeSR™-E8™ (Stemcell technologies) in 6-well plates coated with 0.5% Corning Matrigel® Growth Factor Reduced (GFR) Basement Membrane Matrix, Phenol Red-Free (Cat.NO: 356231) prepared in PBS 1X or advanced DMEM/F12 (Gibco). The cells were grown in incubator at 37C°, 5% CO2, and 5% O2. The ES media was changed daily and when the cells reached 70–80% confluency with healthy pluripotent colonies, they were dissociated into single cells.

2.1.2 hESCs morphology

hESCs represent some distinct and unique characteristics in its ES medium culture (TeSR™-E8™) compared to other adherent cells. It was very crucial to carefully monitor our ES cell morphology during their maintenance and precisely expand the healthy and undifferentiated cells during their passaging. H9 hESCs were checked daily using a light microscope for estimation of their quality and time of passaging. Healthy and high-quality H9 showed approximately round colonies with considerable dense centers, prominent nucleoli, and tightly-packed cells with a high ratio of nucleus:cytoplasm (Figure 4).

For H9 cells over culturing, we considered some notes to monitor them for high quality cells based on recommendation provided by STEMCELL TECHNOLOGIES (<https://www.stemcell.com/assessing-morphology-of-hpscs.html>) as follow:

- a) Approximately up to 3 days after plating, since the colonies were under spreading out and establishing, they usually had looser packing. At this time, some spiky colony edges were observed but they were not abnormal morphology since the colonies increased the dense centers and tightly-packed cells immediately after this time and the typical morphology was achieved around one to two days prior to passaging (Figure 4).
- b) Sometimes, H9 culture on Matrigel showed colonies with slightly asymmetric and merging status, mainly prior to the passaging time but there was no abnormal condition since it occurred when cell aggregates were at low-density seeding and not properly spread (Figure 4).

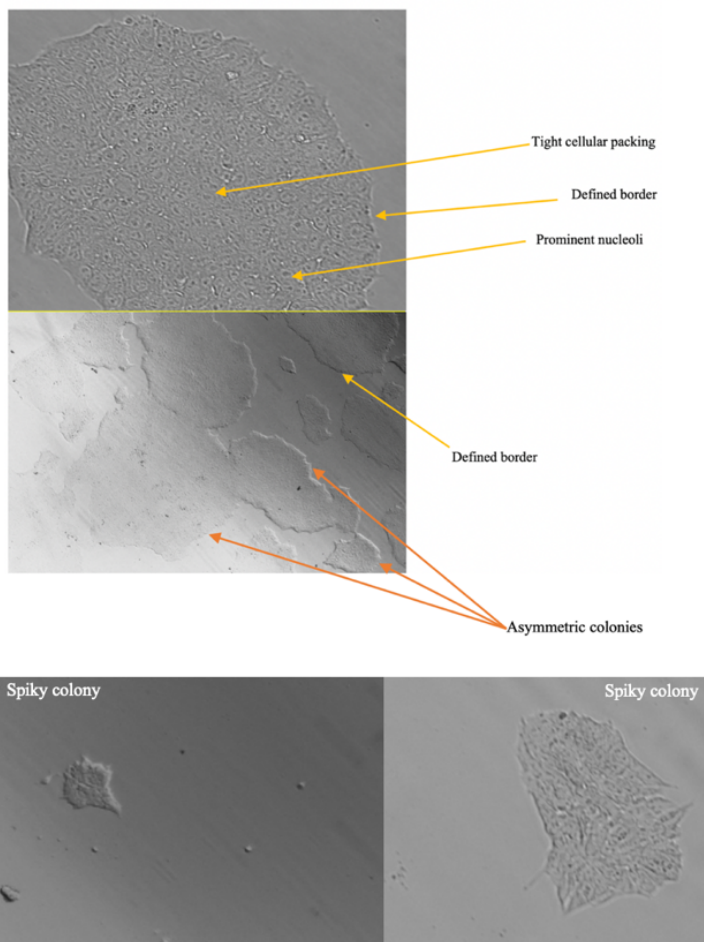


Figure 4. H9 hESCs with normal morphologies.

2.1.3 EDTA-based Passaging of H9 hESCs

We used 0.5 mM EDTA prepared in PBS 1X to passage H9 hESCs approximately every 3 to 4 days when cultures were around 60%-75% confluent in the well, with large and undifferentiated colonies [51].

EDTA-based passaging (described in <https://esibio.com/>) decreased the H9 hESCs death and helped the cells to be rapidly attached on the Matrigel coated plate after seeding them. Usually, we cultured H9 hESCs on 0.5 % Corning Matrigel Matrix prepared in PBS 1X and passaged one well of a 6-well plate of H9 hESCs using 0.5 mM EDTA in PBS 1X for 6 wells (1:6 ratio) or 12 wells of 12-well plates (1:12 ratio) [51].

We performed H9 passaging with EDTA at room temperature (RT) for 4 minutes and only one or two wells each time. Moreover, importantly, we did not dissociate cells into single cells in routine H9 hESCs cultures since it could result in their poor survival and abnormal karyotypes. Usually, we considered small colony aggregates with at least around twenty cells. For EDTA-based passaging protocol we performed following procedure:

1. A 6-well tissue culture plate was coated with 0.5 % Corning Matrigel Matrix prepared in PBS 1X or advanced DMEM/F12.
2. The Matrigel-coated plate was incubated at 37C° for at least 30 min before use.
3. Just prior to the cell dissociation for passaging, the Matrigel solution was removed from the wells and replaced with two mL of TeSR™-E8™ per well of a 6-well plate.
4. The media from the cells ready for passaging was removed.
5. The cell was washed gently with 1 mL of PBS 1X (without Ca²⁺ and Mg²⁺) and then PBS was removed and discarded.
6. One mL of 0.5 mM EDTA prepared in PBS 1X was added to the cells.
7. The culture was incubated at room temperature for 4 minutes.

8. When the cells within colonies represented rounded shape, and started to dissociate, gently the EDTA was removed from the cells.
9. Then, one mL of the pre-warmed TeSR™-E8™ was added to the well by 1000 pipet and pipetted two times to uniformly dissociate cells.
- 10- Around 166 uL of the cell suspension was then dispensed in each well of a new 6-well plate coated with the Matrigel containing 2 mL of TeSR™-E8™.
11. The plate was placed in 37°C incubator at 5% CO₂ and 5% O₂ and it was rocked back-and-forth and side-to-side for uniform distribution of the aggregates throughout the well.
12. The old media was replaced by 2 mL of the fresh RT pre-warmed TeSR™-E8™ daily and cells were monitored daily for their quality and morphology.

2.2. Differentiation into the myogenic lineage

Before starting the differentiation protocol, we considered that cells reached 70–80% confluency and they were undifferentiated. Moreover, the important thing in the protocol was the dissociation of colonies into single cells maintaining their pluripotency and viability, ensuring proper differentiation induction uniformly across all cells.

2.2.1 Culture of H9 hESCs

The cells were grown and maintained in TeSR™-E8™ in 6-well plates coated with Corning Matrigel matrix (prepared in advanced DMEM/F12) in incubator at 37 C°, 5% CO₂, and 5% O₂. When the cells reached 70–80% confluency with healthy pluripotent colonies, they were used for dissociation of colonies into single cells.

2.2.2 Differentiation procedure [4]

We started to differentiate the H9 hESCs in 12-well plate coated with 3% Corning Matrigel. One well of H9 cells in a 6-well plate was used for at least all wells of a 12- well plate (4×10^4 cells/cm of growth area and 1 mL medium/ 1.5×10^5 cells). We started the differentiation procedure with healthy colonies without any signs of differentiated population.

- 1- The old TeSR™-E8™ medium was removed from the cell and washed once with PBS 1X.
- 2- One mL of fresh TeSR™-E8™ medium supplemented with 10 μM Rock inhibitor (ATCC®) was added in each well of 12-well plate. The culture was incubated for 1 hour in a 37 C°, 5% CO₂, and 5% O₂ incubator.
- 3- Then, 1 mL of TeSR™-E8™ was removed and discarded and the cells were washed once with PBS 1X. EDTA dissociation was then performed based on the above-mentioned procedure.
- 4- After EDTA dissociation, the cells were collected using 1 mL of TeSR™-E8™ supplemented with 10 μM Rock inhibitor.
- 5- Total cell number was counted using a hemocytometer. We prepared a cell suspension in TeSR™-E8™ supplemented with 10 μM Rock inhibitor with 1.5x10⁵ cells/mL for seeding as single cells in each well of 3% Matrigel coated 12-well plate. It should be noted that, for paraxial mesoderm induction, cell clusters of 6 to 15 cells after 24 hours of seeding could show better response to GSK3 inhibitor (CHIR99021). Then, the cells in TeSR™-E8™ supplemented with 10 μM Rock inhibitor was incubated overnight in a 37 C°, 5% CO₂, and 5% O₂ incubator.

2.2.3 Differentiation (Day 0–46)

- 6- After 24 hours of incubation (on Day 0, D0), the old media was aspirated and discarded and the cells were gently washed with PBS 1X. The H9 hESCs without process for differentiation was considered as Day 0 (D0).
- 7- On D0, one mL of Essential 6™ Medium (E6, Gibco™) medium supplemented with 10 μM GSK-3 Inhibitor (CHIR99021, Sigma) was added in each well and was incubated 24 hrs. in a 37 C°, 5% CO₂, and 5% O₂ incubator. Usually, few cells are died at Day 1 (D1) of differentiation (less than 20%), representing more efficient mesoderm induction and skeletal myogenesis. On D1, the medium was replaced with E6 medium supplemented with 10 μM CHIR99021 and was incubated 24 hrs. in a 37 C°, 5% CO₂, 5% O₂ incubator. On Day 2 (D2) of differentiation, we assessed the Brachyury (*T*)

gene expression using Quantitative real-time PCR (qRT-PCR) to observe optimal paraxial mesoderm specification.

8- On D2, the old media was removed and discarded, and then the cells were gently washed with PBS 1X. E6 medium without any supplements was then added to the cells. This medium was used from D2 to Day 12 (D12) of differentiation with daily medium change. Using E6, these cells are highly proliferated to give rise to somite stage. On Day 7 (D7), we examined the expression of somite markers *PAX3* (Table 2) using qRT-PCR and immunofluorescence (IF) staining.

9- On D12, the old E6 media was removed and the cells were washed with PBS 1X. Then, one mL of StemPro-34 complete medium (Gibco) supplemented with 5 ng/mL FGF2 (Human recombinant FGF2, Peprotech) was added to the cells. This media was used for up to Day 20 (D20) of differentiation with daily media change. These cells are committed to the myogenic lineage with favored expansion of SMPs.

10- On D20, we followed differentiation protocol for the passaging differentiating cells described by Shelton, M et al. [4] which pre-warmed Collagenase IV was used for disassociation of cells for 20 min in a 37 C°, 5% CO₂, and 5% O₂ incubator and usually we seeded around 3x10⁵ cells in each well of 3% Matrigel coated 12-well plates [4]. Usually, each well of 12-well plate was divided into its three wells and then cells were incubated overnight in a 37 C°, 5% CO₂, and 5% O₂ incubator. After incubation, the old media was removed and discarded and then 1 mL of E6 medium supplemented with 5 ng/mL FGF2 (for both the proliferation and differentiation of SMPs) was added to each well of 3% Matrigel coated 12-well plate for 2 more weeks with daily media change. Cells were re-passaged again each two weeks (Day 35, 50, and so on). Some of the myogenic regulatory factors (MRFs) (*MYOD1*, and *MYOG*) using qRT-PCR and *PAX7* using IF staining were assessed on D30 which should have SMPs [4].

Moreover, we considered our endpoint analysis at D46 and examined the expression of terminally differentiation markers such as MYH (called MYHC and MF20) using IF staining and *MYOG* using qRT-PCR.

2.2.4 IF staining for 8-well chamber slide

- 1- The old media was aspirated and discarded, and then cells were washed 2 times with 200 μ L PBS 1X.
- 2- 150 μ L of 4% paraformaldehyde (PAF) was gently added to each well of an 8 well chamber slide. The cells were fixed in PAF for 15 min at room temperature.
- 3- The PAF was aspirated and discarded.
- 4- The cells were washed 2 times with 200 μ L PBS 1X.
- 5- 150 μ L of 0.3 % triton (Sigma) in PBS 1X was added to each well to permeabilize cell membrane.
- 6- The cells were incubated at RT for 10 min.
- 7- The 0.3 % triton in PBS 1X was aspirated and the cells were washed 2 times with 200 μ L PBS 1X.
- 8- 150 μ L of 1% BSA in PBS 1x was added as blocking solution per well and the cells were incubated at room temperature for 1 hour.
- 9- The blocking solution was aspirated and the cells were washed 2 times with 200 μ L PBS 1X.
- 10- 150 μ L of the primary antibody prepared in blocking solution was added per well and the cells were incubated at 4C^o overnight or 3 hours at room temperature.
- 11- The primary antibody (Table 2) was removed and discarded and the cells were washed 2 times with 200 μ L PBS 1X.
- 12- 150 μ L of secondary antibody prepared in blocking solution was added per well and the plate was covered in foil and incubated at room temperature for 45 min. All antibodies used in this study are given in Table 2.
- 13- The secondary antibody was aspirated and discarded and the cells was gently washed 2 times with 200 μ L PBS 1X.
- 14- 100 μ L of DAPI solution (1 μ g/mL⁻¹) was added to each well and the plate was covered by foil and incubated at room temperature for 5 min.
- 15- DAPI solution was removed and discarded and the cells were washed two times with 200 μ L PBS 1X and then 200 μ L PBS 1X was added on the cells and observed under a fluorescent microscope.

Table 2. Antibodies used in our study for confirmation of muscle stage markers

Antibody name	Company	Cat No.
PAX3	Proteintech	21386-1-AP
PAX7	Proteintech	20570-1-AP
Myosin Heavy Chain, Fast isoform	Sigmaaldrich	M4276
Myosin Heavy Chain, Slow isoform	Sigmaaldrich	M8421
MF20, mouse	DSHB, Developmental Studies Hybridoma Bank	AB_2147781
Anti-rabbit Alexa Fluor 568, goat	Invitrogen	A11011
Anti-mouse Alexa Fluor 488, goat	Invitrogen	A11001
DAPI, FluoroPure™ grade	Invitrogen	D21490

2.2.5 QRT-PCR analysis

2.2.5.1 RNA extraction by Qiazol Reagent

- 1- Growth media was removed and the cells were washed with PBS 1X.
- 2- 400 μ L of QiAzol Reagent (Qiagen) per one well of 12-well plate was added directly to the culture.
3. The lysate was pipetted up and down several times to homogenize it.
- 4- The homogenized lysate was incubated for 5 minutes.
- 5- 100 μ L of chloroform was added and then the sample was incubated for 3 minutes.
- 6- The sample was centrifuged for 15 minutes at $14,000 \times$ RPM at 4°C .
- 7-The colorless upper aqueous phase containing the RNA was transferred to a new 1.5 ml tube.
- 8- RNase-free glycogen was then added to the aqueous phase.
- 9- 250 μ L of isopropanol was added to the aqueous phase.
- 10-The sample was then incubated for 10 minutes.

11-Then, the sample was centrifuged for 10 minutes at $14,000 \times RPM$ at $4^{\circ}C$.

12-The supernatant was removed and discarded.

13- The pellet was then washed two times with 250 μL of 75% ethanol with centrifugation for 5 minutes at $14,000 \times RPM$ at $4^{\circ}C$.

14- The supernatant was discarded and the RNA pellet was air dried.

15-The pellet was then resuspended in 20–50 μL of RNase-free water and was incubated in a heat block at $50^{\circ}C$ for 10 minutes.

2.2.5.2 QRT-PCR

The sample quality control was determined using Agilent Bioanalyzer 2100. The samples had an RNA integrity number ranging from 8 to 10. mRNA quantitation was carried out by qRT-PCR using the The SensiFAST SYBR® No-ROX, One-step Kit (Bioline, meridian|BIOSCIENCE) according to manufacture' instructions and normalized on *PUM1* mRNA. All oligonucleotide sequences used for qRT-PCR for muscle differentiation stages are given in Table 3.

Table 3. All oligonucleotide sequences used for qRT-PCR for confirmation of muscle differentiation stage markers.

Oligo name	Oligo Sequence
PUM1-F	GGGCATGGAGCCTCTTCAGTTT
PUM1-R	GGACAGCAAGCGCATTAGGTCTTT
F-TBXT-QPCR	CCAGATCATGCTGAACTC
R-TBXT-QPCR	CTGTGATCTCCTCGTTCT
F-PAX3-QPCR	AAGATCCTGTGCAGGTAC
R-PAX3-QPCR	CTGATGGAACCTCACTGAC
F-MYOG-QPCR2	ATGGAGCTGTATGAGACATC
R-MYOG-QPCR2	ACACCGACTTCCTCTTACA
F-MYOG.QPCR3	GCTGTATGAGACATCCCCCTA
R-MYOG.QPCR3	CGACTTCCTCTTACACACCTTAC
F-MYHC-QPCR2	GTACCTCCGCAAGTCAGAG
R-MYHC-QPCR2	TCCCATCTCGGTTTCAG

2.2.6 RNA sequencing library preparation for transcriptome sequencing

The RNA sequencing library was prepared using Illumina TruSeq RNA Sample Prep Kit v2. We used 1 µg of total RNA with high quality determined by Agilent Bioanalyzer 2100 (RNA integrity number ranging from 8 to 10). In overall, at the first step, the poly (A) mRNAs were purified using oligo-dT attached magnetic beads. In the next step, the mRNA was fragmented into small fragments using divalent cations at high temperature. Then, with the use of reverse transcriptase and random oligos, the first strand cDNA was produced from the fragmented mRNA. Next, the second strand cDNA was synthesized using DNA Polymerase I and RNase H. End repairing was carried out for these cDNA fragments, and the single 'A' base addition step was then performed. Next, the adapters ligation was carried out on these fragments. At the final steps, the products were purified using magnetic beads, enriched with PCR, and it was purified using magnetic beads to make the desired and optimal cDNA library.

2.2.7 RNA Sequencing Analysis

Sequencing was performed on the Illumina NextSeq 500 platform. After quality controls with FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 25 November 2021)), raw reads were aligned to the GRCh38 human genome reference (hg38) using STAR 2.7.1a [52] (with parameters `-outFilterMismatchNmax 999 -outFilterMismatchNoverLmax 0.04`). Gene expression levels were quantified with featureCounts v1.6.3 [53] (options: `-t exon -g gene_name`) using GENCODE basic gene annotation release 27. Multi-mapped reads were excluded from quantification. Gene expression counts were next analyzed using the edgeR package [54]. After low expressed genes filtration (1 count per million (CPM) in less than 3 samples), normalization factors were calculated using the trimmed-mean of M-values (TMM) method (implemented in the `calcNormFactors` function) and CPM were obtained using normalized library sizes. ANOVA-like test for differential expressed genes in H9 cells was carried out by fitting a Generalized Linear Model (GLM) to all time points and performing Quasi-Likelihood F-test (QLF) (the design matrix formula

= “~time” sets up time 0 as reference point). Genes with False Discovery Rate (FDR) less than or equal to 0.001 and $|\log FC| > 1$ were considered as differentially expressed. K-means clustering of gene expression profiles was performed on differentially expressed genes only, fixing 9 as cluster number. CPM values were scaled as Z-scores across samples before computing distances. Gene expression heatmaps were generated using the ComplexHeatmap R package[55]. Gene ontology and pathways enrichment of clusters were performed using ClusterProfiler R package [56].

2.3. Expression of TCGA-breast adenocarcinoma (BRCA) genes affecting OS across muscle cell differentiation

As mentioned in introduction section, pectoralis muscle shows altered expression of genes involved in muscle protein homeostasis [35, 57]. Based on these evidences, using an R/Bioconductor package, GDCRNATool (Li et al., 2018)[58, 59], we provided genes with altered expression affecting OS from The Cancer Genome Atlas (TCGA, <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>.) breast cancer (BRCA) and then investigated across our muscle differentiation stages, help identify their corresponding pathways. We also used Enrichr (<https://maayanlab.cloud/Enrichr/>) [60] to look for gene ontologies (GO term : Biological Process, BP) of the uncharacterized genes found in TCGA-BRCA.

In this part of study, firstly, we used GDCRNATool to extract, arrange, and analyze differential expression of lncRNAs and mRNAs along with clinical data of breast cancer patient from TCGA. This database has raw data of clinical, genomic, epigenomic, transcriptomic, and proteomic studies of 33 tumor types. We analyzed TCGA-BRCA tumor tissue data compared to their matched adjacent non-cancerous tissues. The appropriate data of these breast tumors and their matched adjacent non-tumor tissues were extracted. We started our analysis with 1222 mRNA sequencing data and 1097 clinical data. Next, using GDCRNATool, we studied univariate survival analysis with `gdcSurvivalAnalysis` function considering Kaplan Meier (KM) analysis. In this analysis, patient data were grouped into high and low-expression according to the median. Then, among genes with altered

expression, we only selected those were significantly associated with OS in BRCA patients and looked for their involvement in muscle differentiation stages [58, 59]. Finally, we only selected those genes with downregulation as possible novel tumor suppressor genes for our further analysis, including gene set enrichment analysis (GSEA) using Enrichr.

Briefly, in this package, RNA sequencing and clinical data were downloaded from TCGA. Next, metadata of RNA sequencing were parsed using the `gdcParseMetadata` function to efficiently organize the RNA and clinical data. Moreover, using the `gdcFilterDuplicate` and `gdcFilterSampleType`, duplicated samples and samples without primary tumor and normal solid tissue, respectively, were excluded from the study. Then, each data of RNA sequencing and its matched clinical data were separately merged. In next step, normalization of RNA sequencing data was carried out by `gdcVoomNormalization` function applying the TMM method in `edgeR` and the `voom` method in `limma`. We excluded genes with low-expression, `logcpm` less than one, from our TCGA-BRCA analysis. Furthermore, `gdcDEAnalysis` and the `gdcDEReport` functions in `GDCRNATool` helped us to provide all differential gene expression and distinct differentiation expression of each lncRNA and mRNA, respectively [58, 59]. The `gdcDEAnalysis` function, which provides `limma`, `edgeR`, and `DESeq2`, was applied to find differentially expressed genes (DEGs) between TCGA-BRCA tumors and their matched normal counterparts [58, 59].

2.4 Investigation of ubiquitin proteins, mainly *UBE2C*, in muscle differentiation stages

We explained that pectoralis muscle of breast cancer patients shows an elevated expression of genes involved ubiquitin-mediated proteolysis. Therefore, based on these data, in another part of our study, we looked for the expression of ubiquitin proteins (Table 4 [61] and 5), mainly the main member of ubiquitin proteins, *UBE2C* which is overexpressed in all TCGA cancers, and its co-expressed genes, across muscle differentiation. This may help find their role and pathways in muscle differentiation stages and also in muscle tissue-related cancers, including breast cancer. Since *UBE2C* is a key regulator of cell cycle progression and together with *APC/C* is involved in the initiation of ubiquitin

chain formation on APC/C substrates in the ubiquitin-proteasome pathway [49, 50] and also is contributed to the tumorigenesis of a several of cancers, mainly breast adenocarcinoma [36, 40], we mainly focused on this gene and its co-expressed genes in cancer and muscle differentiation stages.

Table 4. Ubiquitin E2 enzymes [61]

Ubiquitin E2 enzymes	
UBE2A, UBE2B, UBE2C	UBE2V1, 2
UBE2D1, 2, 3, 4	UBE2W
UBE2E1, 2, 3	BIRC6
UBE2G1, 2	UBE2F
UBE2H	UBE2I
UBE2J1,2	UBE2L6
UBE2K	UBE2M
UBE2L3	UBE2Z
UBE2N, UBE2NL	ATG10
UBE2O	ATG3
UBE2Q1, 2, L	
UBE2R1, 2	
UBE2S	
UBE2T	
UBE2U	

Table 5. Ubiquitin Ligases*

Gene name	Gene name	Gene name	Gene name
AMFR	SCF/ β -TrCP	RNF5	Parkin
APC/Cdc20	SCF/FBW7	RNF8	PELI1
APC/Cdh1	SCF/Skp2	RNF19	Pirh2
C6orf157	SHPRH	RNF190	PJA1
Cbl	SIAH1	RNF20	PJA2
CBLL1	SIAH2	RNF34	RFFL
CHFR	SMURF1	RNF40	RFWD2
CHIP	SMURF2	RNF125	Rictor
DTL (Cdt2)	TOPORS	RNF128	MARCH-I
E6-AP	TRAF6	RNF138	MARCH-II
HACE1	TRAF7	RNF168	MARCH-III
HECTD1	TRIM63	MEKK1	MARCH-IV
HECTD2	UBE3B	MIB1	MARCH-VI
HECTD3	UBE3C	MIB2	MARCH-VII
HECW1	UBR1	MycBP2	MARCH-VIII
HECW2	UBR2	NEDD4	MARCH-X
HERC2	UHRF2	NEDD4L	ZNRF1
HERC3	VHL	mahogunin	
HERC4	HYD	WWP1	
HERC5	ITCH	WWP2	
HUWE1	LNX1		

* The list of these enzymes was extracted from <https://www.cellsignal.com/learn-and-support/reference-tables/ubiquitin-ligase-table#anc1>.

Co-expressed genes with *UBE2C* from TCGA cancers were selected based on the following criteria to investigate them across muscle differentiation stages (all analyses related to this part of study were published in our previous study [36]): 1-Genes with the positive *UBE2C* expression correlations with strong to very strong in all 27 cancers (*MYBL2*, *TROAP*, *CDC20*, *CENPA*, *KIFC1*, *CDK1*, *KIF4A*, and *KIF20A*) [36], 2-Genes with products with TF binding sites on both the promoter and enhancer regions of *UBE2C* which showed positive correlations with *UBE2C* in all 27 cancers (*FOXMI*, *E2F1*, and *RAD51*) [36], 3-Tumor suppresser genes with the positive expression correlation with *UBE2C* in almost all cancers (*BUB1B*) and one with down-regulation in the most of the 27 cancers [36], 4-*UBE2S* which elongates and branches the ubiquitins added by *UBE2C* for formation of Lys-11 (K11)-linked polyubiquitination on its substrates [36]. Among these genes *MYBL2*, *FOXMI*, *E2F1* have TF function, *CDK1* and *BRCAl* act as Epifactors and *RAD51* has role as TF and Epifactor.

Chapter 3

RESULTS

3.1 Confirmation of H9 differentiation to skeletal myogenic progenitors and skeletal muscle cells using stage markers

24 hours after seeding the single cells, we revealed cell clusters with 6-15 cells each (Figure 5A), representing optimal cell density and cluster for mesoderm induction. On D1 of differentiation treated with 10 μ M CHIR99021, H9 cells showed cell death but less than 25% (Figure 5B), representing efficient mesoderm induction. On D2 of differentiation, we examined the Brachyury (*T*) gene expression using qRT-PCR and found its significant high expression compared to D0 (stem cell), indicating the efficient mesoderm specification (Figure 6A).

From D6-12 we considered for the evaluation of *PAX3* as somite marker. Therefore, on D7, we found higher expression of somite markers *PAX3* using qRT-PCR and immunofluorescence (IF) staining (Figure. 6B). On D20, we observed the expected highly proliferative myogenic lineage cells that needed to be passaged (Figure 7A). On D20, we assessed the expression of *PAX7* using IF and it showed its higher expression compared to D0 (stem cell) (Figure 7B), indicating cell commitment to the myogenic lineage with preferential expansion of SMPs cells. On D30, we examined the expression of *MYOG* using qRT-PCR (Figure 7C) and the results showed its significant expression, indicative of progressing to the early dermomyotome-like stage. On D30 upward we observed the sign of myocytes with more obvious on D40 upward (Figure 8A). On D46 we looked for the terminal muscle differentiated marker (MF20) using IF which showed its expression in this stage (Figure 8B) and also some fused cells (Figure 9A), confirming the efficiency of the muscle differentiation processes. Moreover, to verify the presence of both SMPs expressing *PAX7* and also terminal muscle stages, we merged the expression of *PAX7* and *MYH3* (MF20), which tested positive for both with higher rates for MF20, confirming the presence of SMPs for muscle cell-based therapy (Figure 9B).

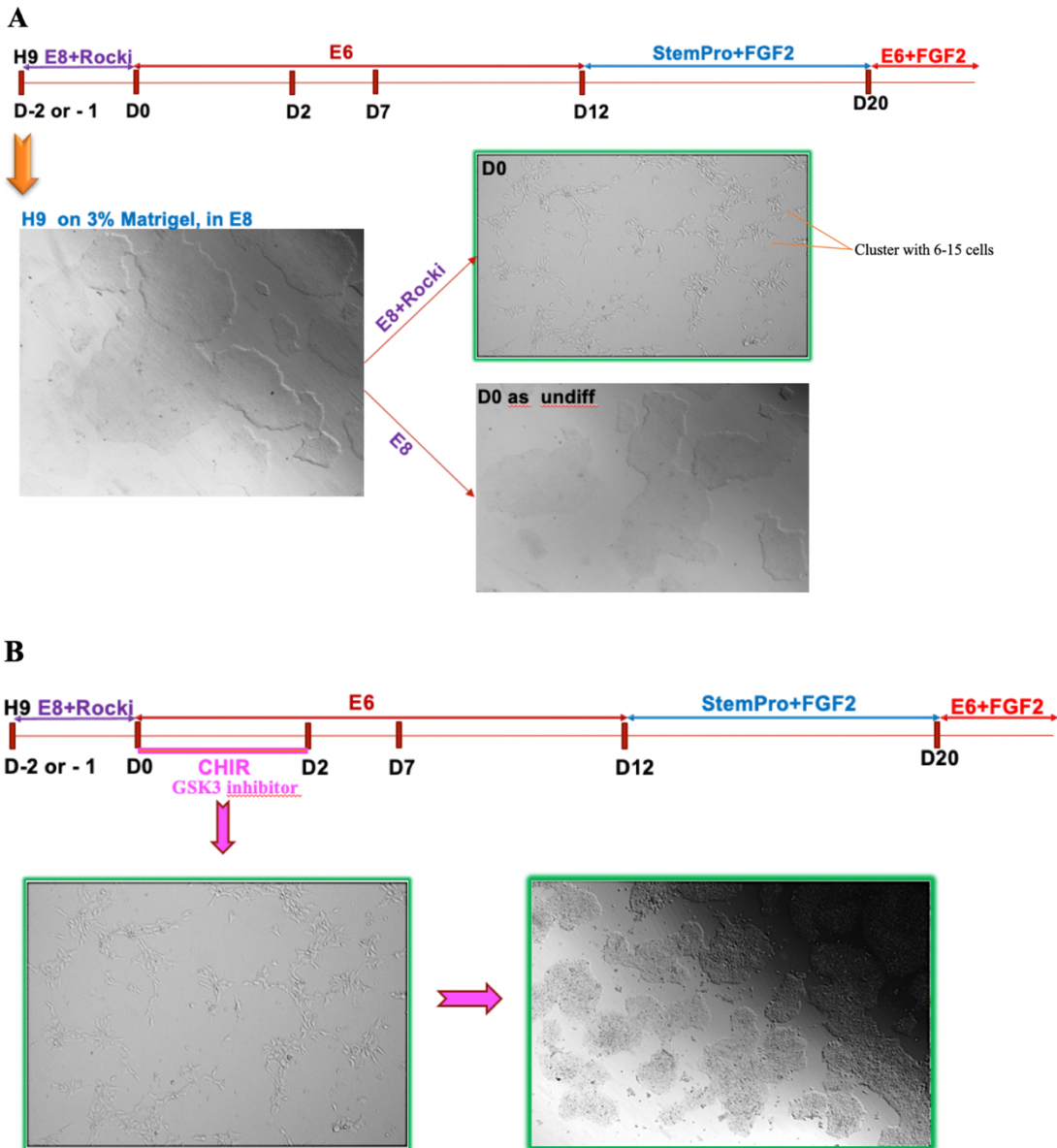


Figure 5. A. Optimal H9 hESCs to be used for preparation of single cell suspension (Left) and H9 cells dissociated into singles cells with clusters of 6-15 cells after overnight incubation in E8 supplemented with Rock inhibitor (right, upper). **B.** CHIR99021 treatment on H9 cells with clusters of 6-15 cells on average. Cells on the left side are D0 in E8 medium supplemented with Rock inhibitor and cells on the right side are D1 after 24 hrs. in E6 supplemented with CHIR, representing signs of cell death due to CHIR induction.

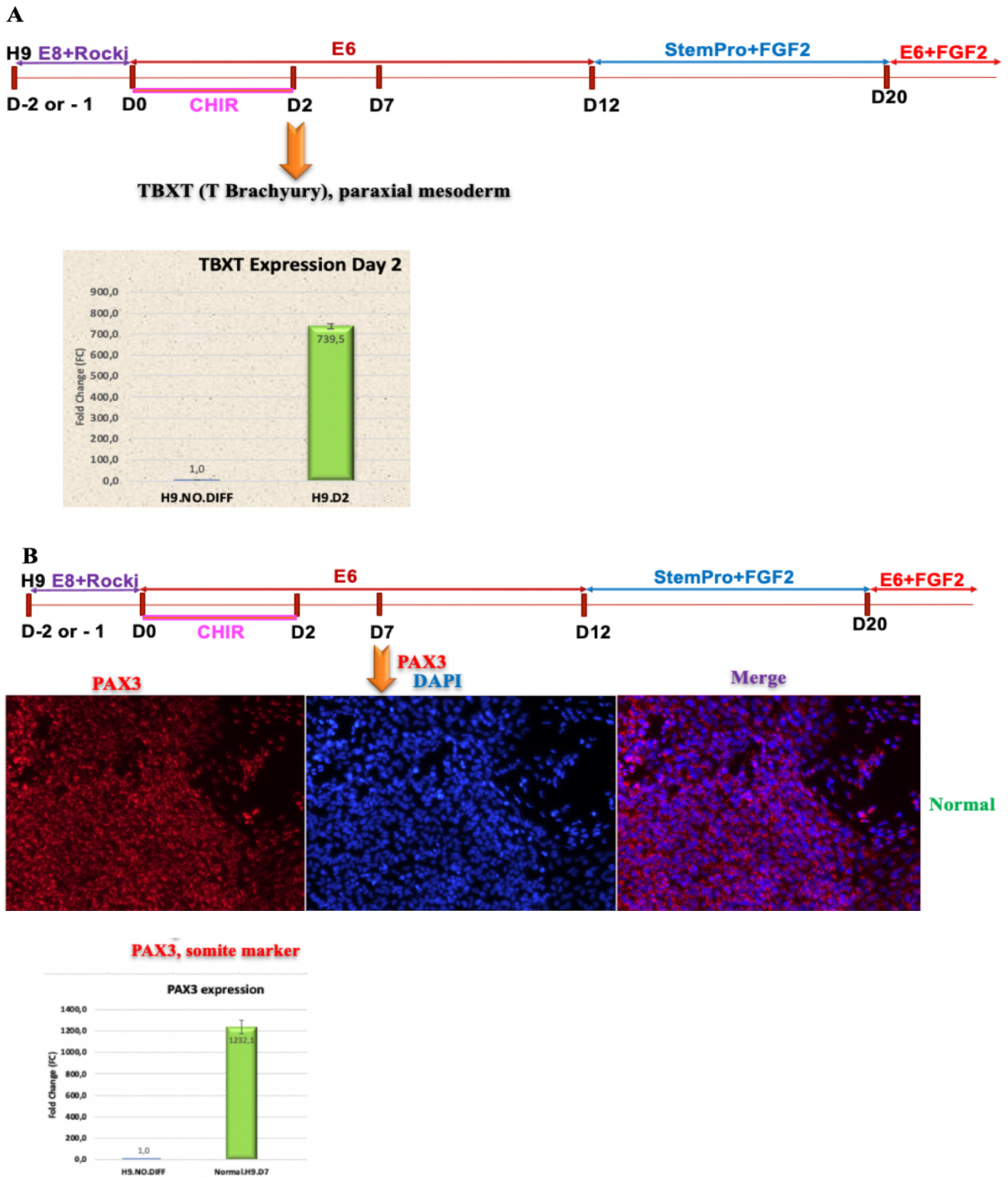


Figure 6. A. Significant higher expression of the Brachyury (*T*) using qRT-PCR at D2 of differentiation. B. Higher expression of PAX3 using qRT-PCR and IF staining at D7 of differentiation. Nuclei were visualized with DAPI ($1 \mu\text{g mL}^{-1}$).

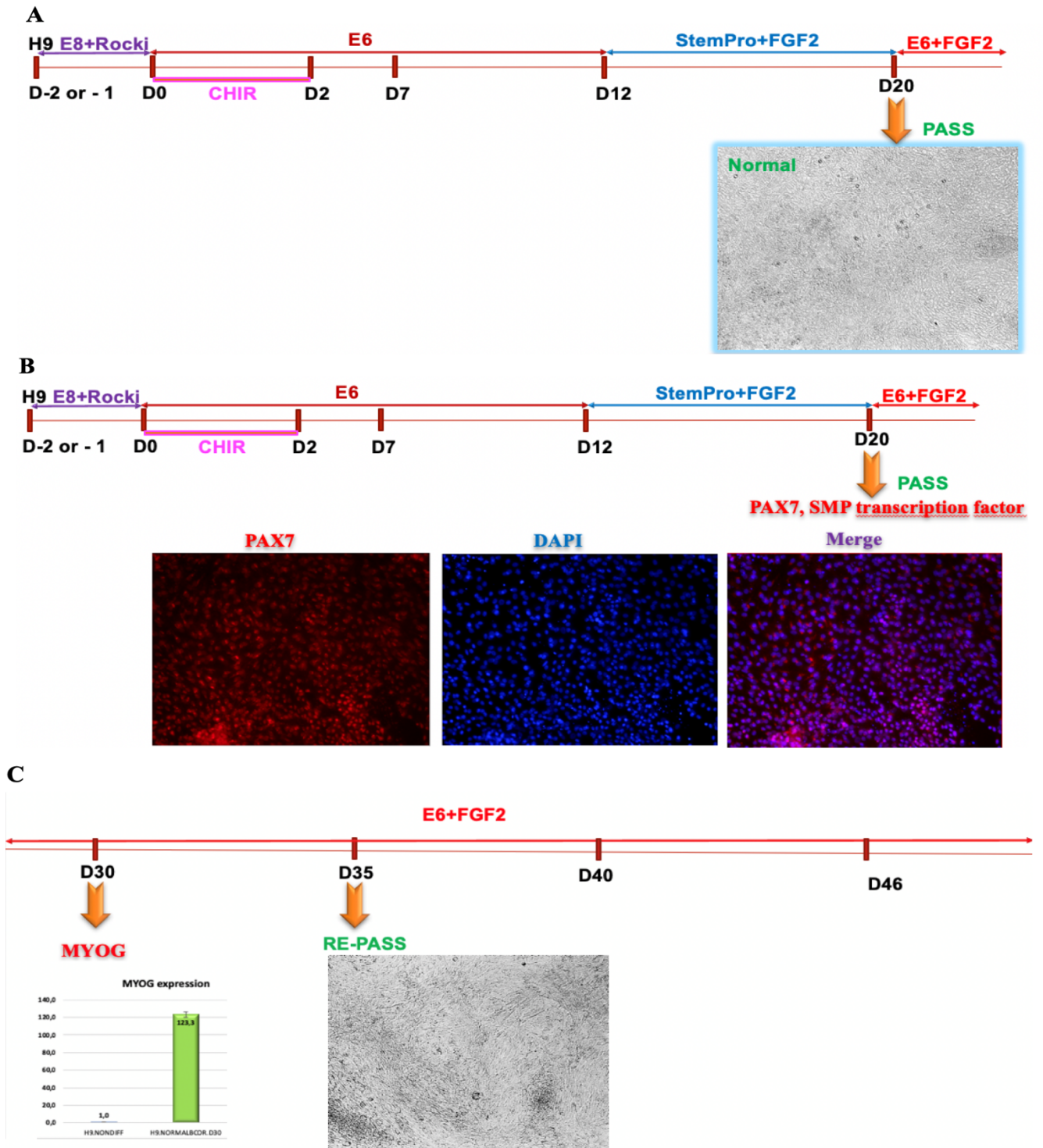


Figure 7. **A.** High rate of proliferated myogenic lineage cells ready for passaging at D20. **7B.** High expression of PAX7 using IF staining at D20 of differentiation, presence of SMPs and myogenic lineage. **7C.** High expression of *MYOG* using qRT-PCR at D30 of differentiation. Nuclei were visualized with DAPI ($1 \mu\text{g mL}^{-1}$).

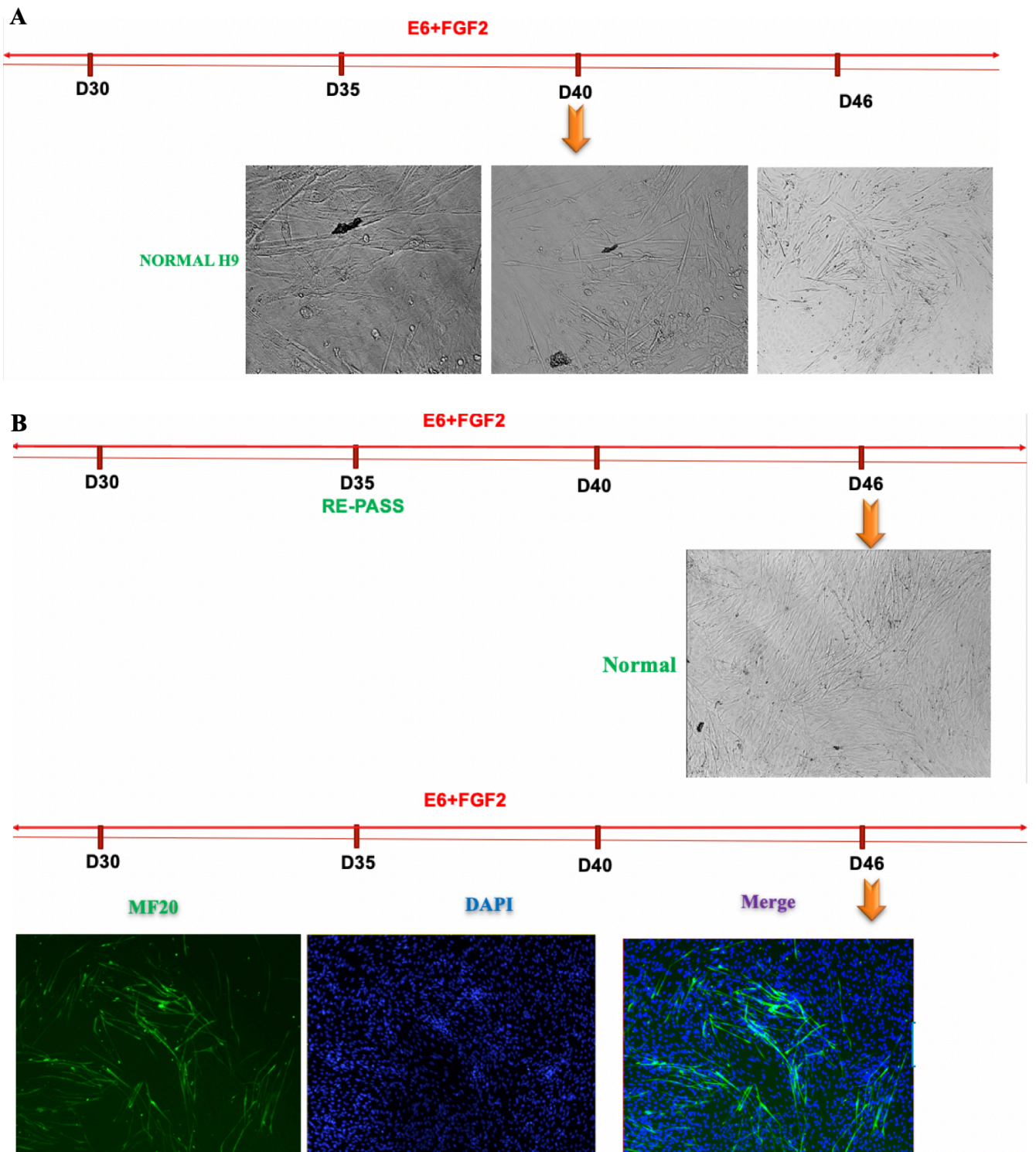


Figure 8 A. Presence of myocytes on Day 40. B. Presence of more myocytes and high expression of the terminal muscle differentiated marker, MF20 using IF staining. Nuclei were visualized with DAPI ($1 \mu\text{g mL}^{-1}$).

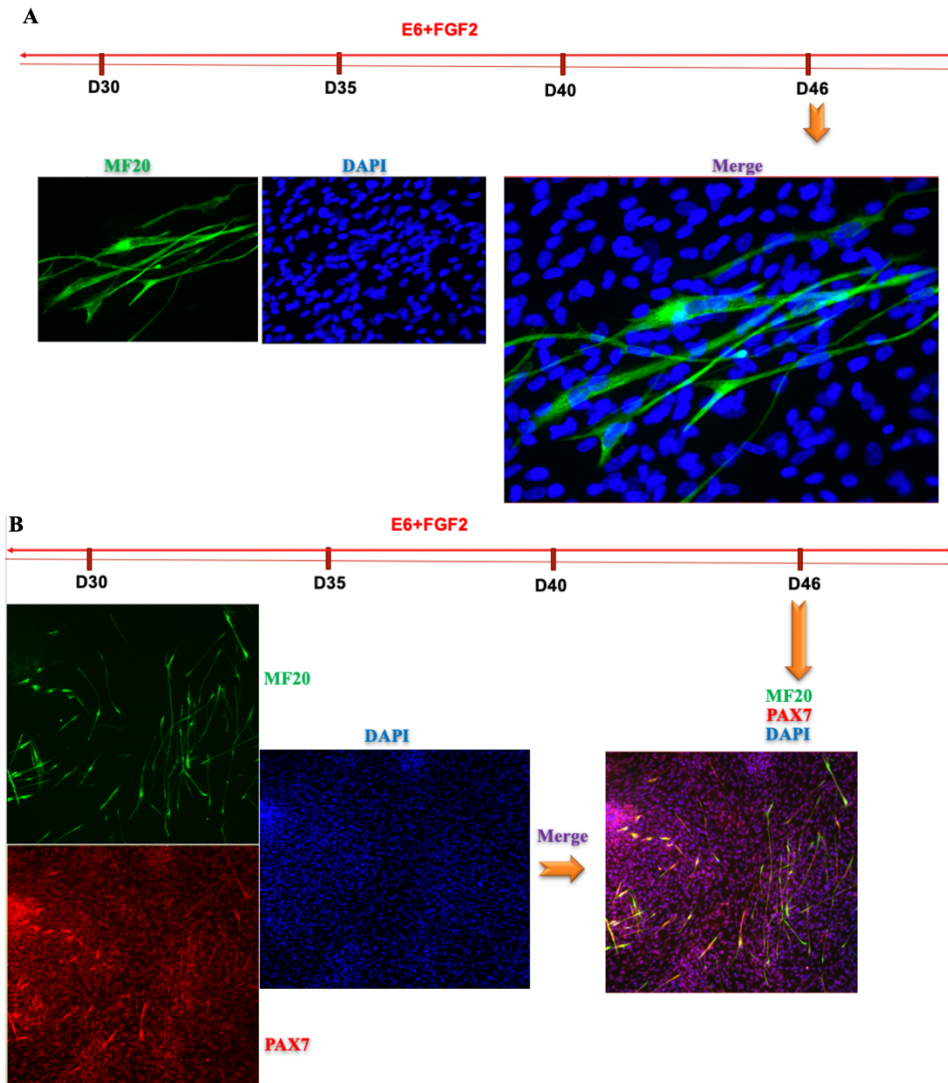


Figure 9. A. Presence of some fused myocytes in IF staining by terminal muscle differentiated marker, MF20.
B. IF staining for both PAX7 (intense red) and MF20 (green) which shows the presence of PAX7+expressing SMPs reside between terminally differentiated cells. Nuclei were visualized with DAPI ($1 \mu\text{g mL}^{-1}$).

3.2 Transcriptome profiling of muscle differentiation stages

3.2.1 RNA sequencing quality control and Principal Component Analysis (PCA)

Our RNA sequencing analysis results showed the optimal read alignments and sequencing quality (Figure 10).

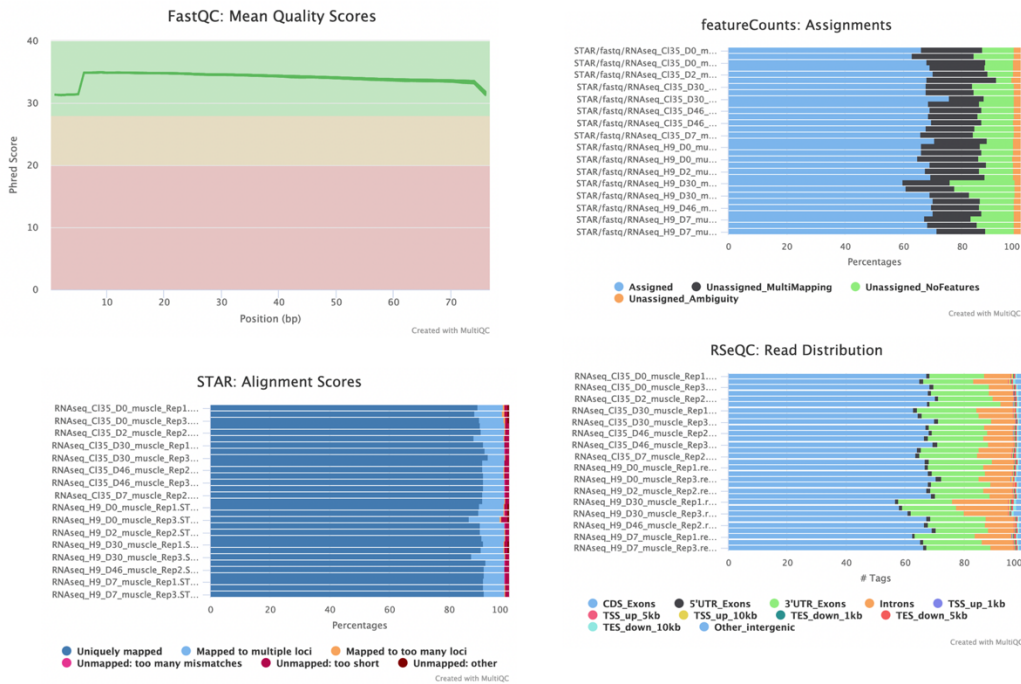


Figure 10. General RNA-sequencing quality controls such as Mean Quality Score, Assignment, Alignments Scores, and Read Distribution.

Principal Component Analysis (PCA), showed that cells in each stage were in the same group but separated according to differentiation stage, indicating robust efficiency of the protocol and good reproducibility during H9 hESCs differentiation to muscle cells at D0, D2, D7, D30, and D46 (Figure 11).

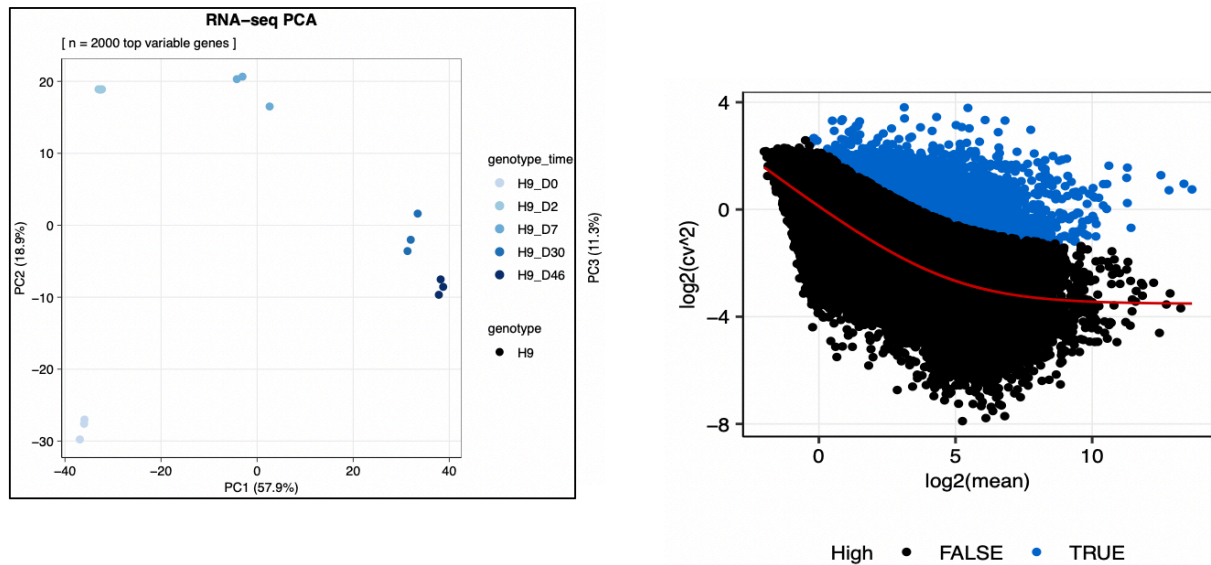


Figure 11. PCA of RNA sequencing samples for H9 hESCs differentiation to muscle cells at 5-time course (D0, D2, D7, D30, and D46). The color intensity represents each differentiation stage, with the lowest intensity for D0. Top 2000 most variable genes were considered for PCA.

In addition to the PCA for sample distribution, we also provided loading plots from PCA for genes extracted from the RNA sequencing data. Observing the data from different angles enabled us to find important genes (Figure 12) and then find the GO:BP and MF and also KEGG pathways related to these genes (Figure 12). For example, for GO:BP the most significant terms were in PC1, PC2, and few terms in PC3. For PC1 of GO:BP, the terms were multicellular organismal process, extracellular matrix organization, movement of cell or subcellular component, cell migration, and cell surface receptor signaling pathway (Figure 12). Regarding PC2 of GO:BP, it revealed epithelium development, system development, regulation of multicellular organismal development, cell-cell signaling, and cell adhesion (Figure 12). Its PC3 showed involvement of multicellular organismal

process, cell adhesion, anatomical structure formation involved in morphogenesis, cell differentiation, and regulation of multicellular organismal process (Figure 12). In relation to the GO:MF, the main terms were enriched in PC1 which included extracellular matrix structure constituent, extracellular matrix binding, growth factor binding, calcium ion binding, and collagen binding. Regarding the KEGG pathways, the main pathways for these genes were in PC1 which were as follow: ECM-receptor interaction, protein digestion and absorption, focal adhesion, PI3K-Akt signaling pathway, and hypertrophic cardiomyopathy (HCM) (Figure 12). However, the main GO:BP and KEGG pathways analysis will be provided for each stage/s in following sections.

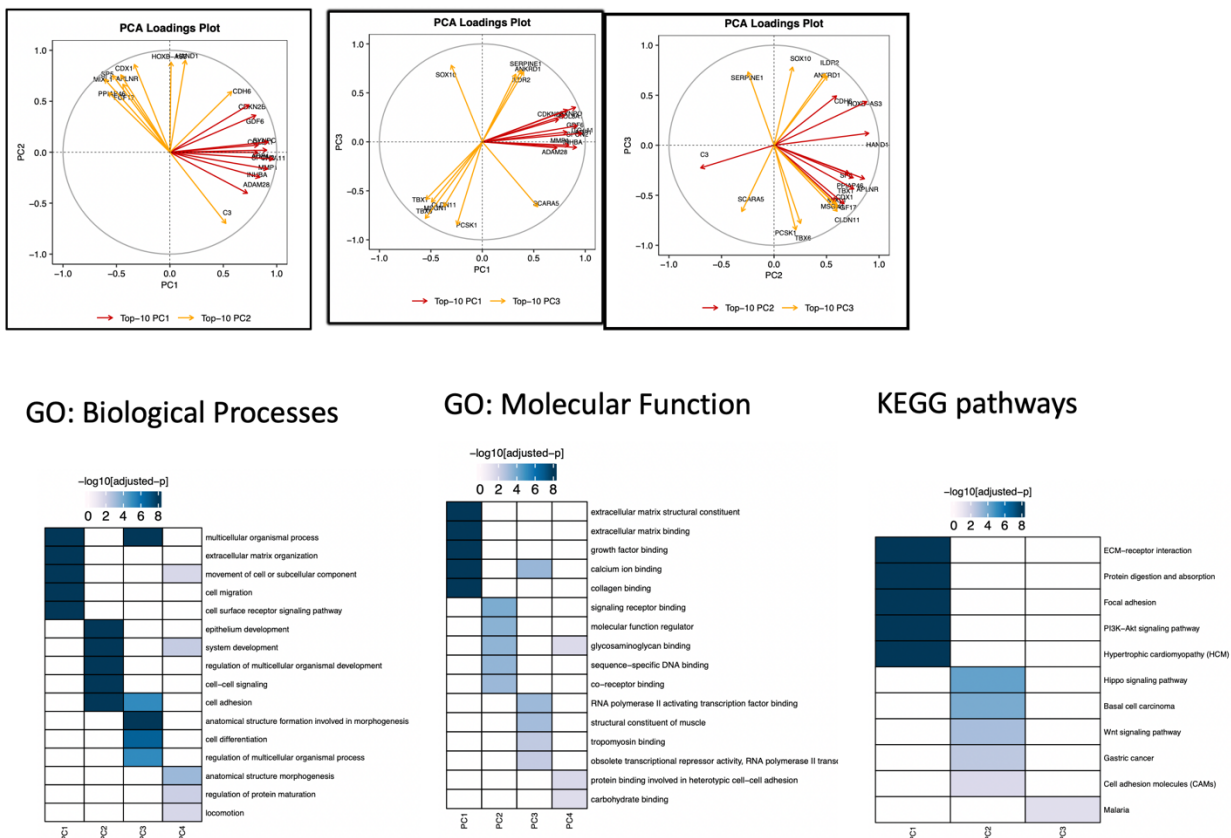


Figure 12. PCA for genes extracted from RNA sequencing data and their GO:BP and GO:MF and KEGG pathways.

3.2.2 Confirmation of stage specific markers

Our transcriptome profiling results confirmed the expected behavior for the regulation of stage-specific markers as shown in Figure 13. In stem cell stage (D0), cells showed high expression of embryonic stem cell markers *POU5F1*, *SOX2*, and *NANOG*. D2 revealed the high expression of expected paraxial mesoderm stage markers, *TBXT*, *MSGN1*, and *TBX6*, in order. On D7, we observed expression of *PAX3*, *MEOX1*, *MEOX2*, and also *PAX7*, representing an efficient somite stage. On D30, we looked for SMPs markers and data revealed the high expression of *MYOG*, *MYOD1*, and *PAX7*, in order. Finally, at terminal stage, the expected markers of *MYH7* and *MYOG* (also *PAX7* for presence of SMPs in this protocol) showed high expression compared to D0, confirming the efficient muscle cell generation with expression of both terminal stage markers and *PAX7*⁺ expressing SMPs. It should be noted that at D7, the expression of *PAX7* also was higher along with *PAX3* which is expected [6, 62, 63]. Moreover, at D7 we also observed higher expression of *SOX2*, which is expected for somite stage [64].

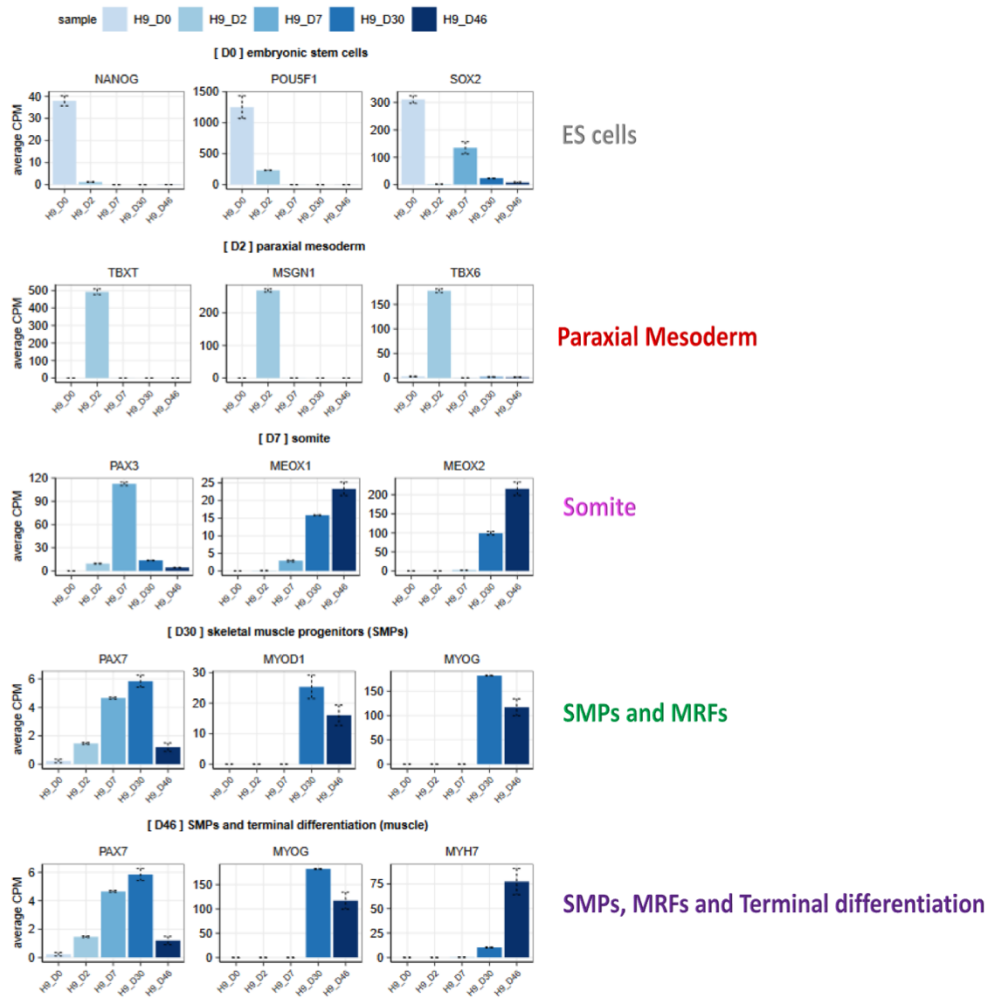


Figure 13. Expected pattern of specific markers of muscle differentiation stages from H9 hESCs. Data were extracted from RNA sequencing and average CPM were considered for comparison of gene expression between each stage.

3.2.3 Differential gene expression and time-course gene expression profiling

Our differential gene expression analysis identified 8641 genes across the differentiation stages. The results of heatmap showed 9 clusters of gene expression which clearly displayed the distinct patterns of regulation over the time course (Figure 14). The clusters were grouped in stem stage (cluster 8), early stage (clusters 2, 4, and 5), middle stage (clusters 1 and 7), late stage (cluster 9) and late & stem stage (clusters 3 and 6) based on the gene expression in clusters (Figure 14).

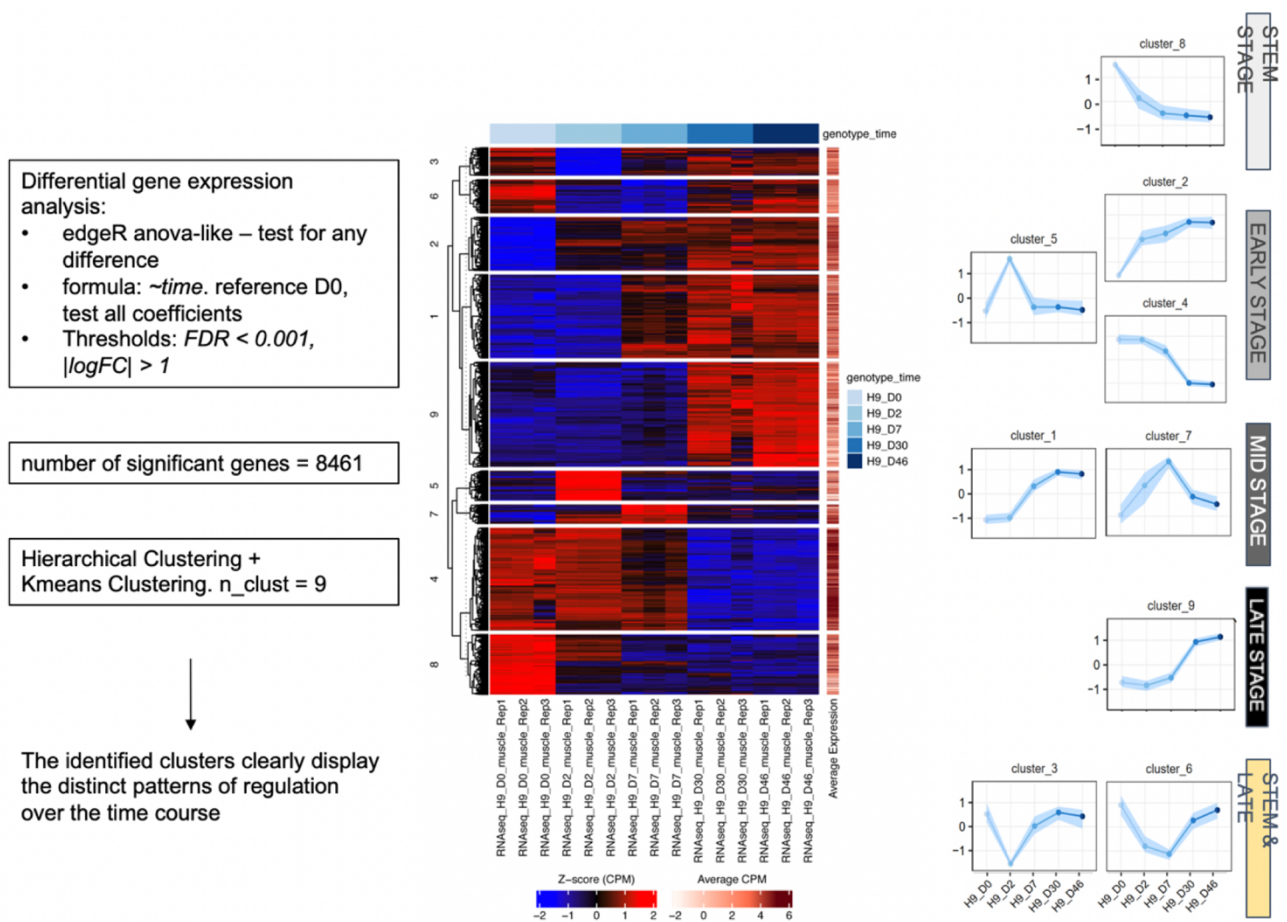


Figure 14. Time-course gene expression profiling and gene clusters of H9 hESCs differentiation to muscle cells.

3.2.4 Cluster profiling based on GO: Biological Processes and KEGG pathways

Cluster profiling for GO: Biological Process (BP) term revealed different processes for each of these 5 stages (D0, D2, D7, D30, and D46) (Figure 15). For example, GO: BP with muscle cell development was specific for the late stage (Figure 15), GO:BP term of somite development was observed for early and middle stage clusters. GO: BP terms of muscle system process and muscle organ development were found from early stage clusters to late stage clusters (it was confirmed by expressing *PAX7* from D7) (Figure 15). Also, GO: BP term of muscle organ development was identified for both stem and late stage clusters (clusters 3 and 6). GO: BP term of mesenchymal development was specific for early and middle stage clusters. Moreover, GO: BP term of connective tissue development was observed for early, middle, and late stage clusters. Furthermore, GO: BP term of embryonic skeletal system development was seen in early and middle stage clusters (Figure 15). Other BP terms are given in figure 15.

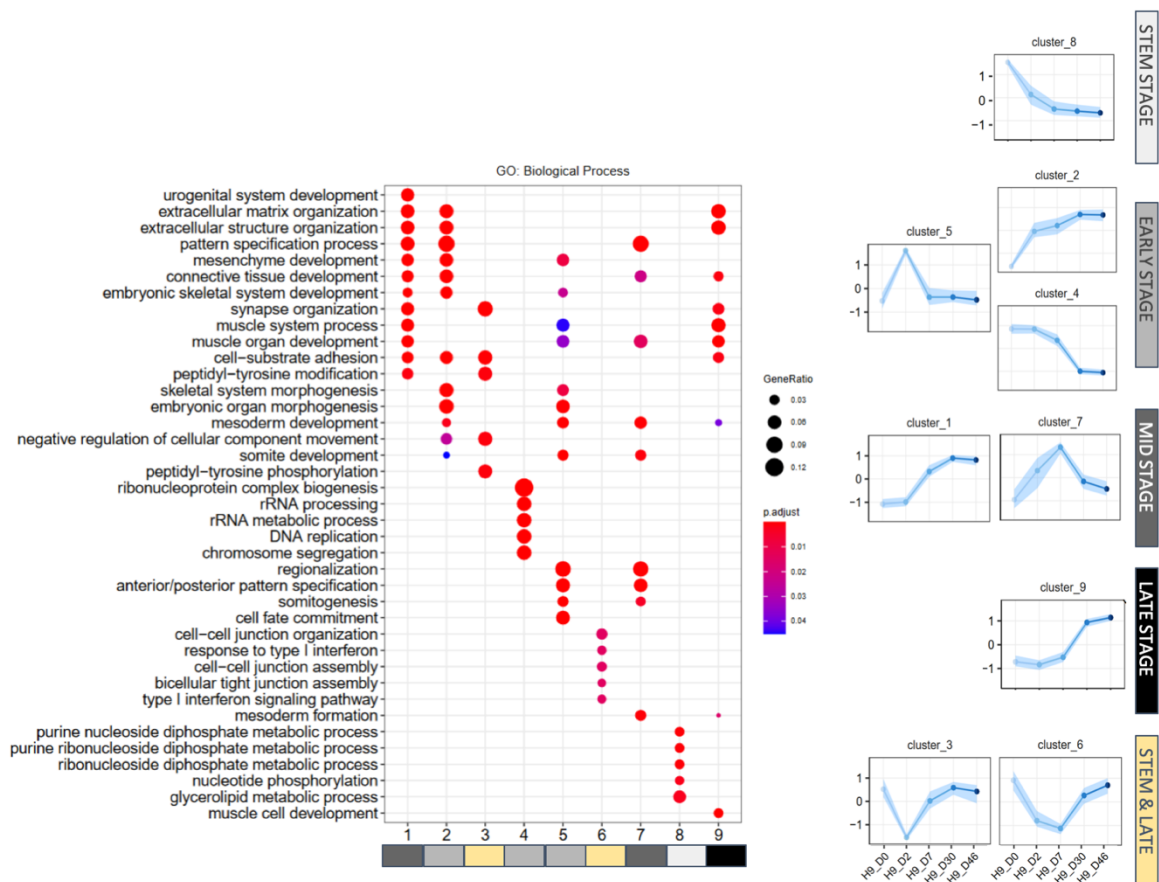


Figure 15. Cluster profiling based on GO: BP terms during H9 hESCs differentiation to muscle cells.

Regarding KEGG pathways, we found different key pathways for each cluster or common between them (Figure 16). For example, PI3K-Akt signaling pathway and Rap1 signaling pathways were specific for middle stage clusters and stem & late stage clusters. Moreover, signaling pathways regulating pluripotency of stem cells was specific for stem stage and early stage clusters. Furthermore, we found Wnt signaling pathway specific for early stage, middle stage, and stem & late stage clusters. These important pathways in muscle differentiation are also among the main cancer pathways (Figure 16). Other KEGG pathways are given in figure 16.

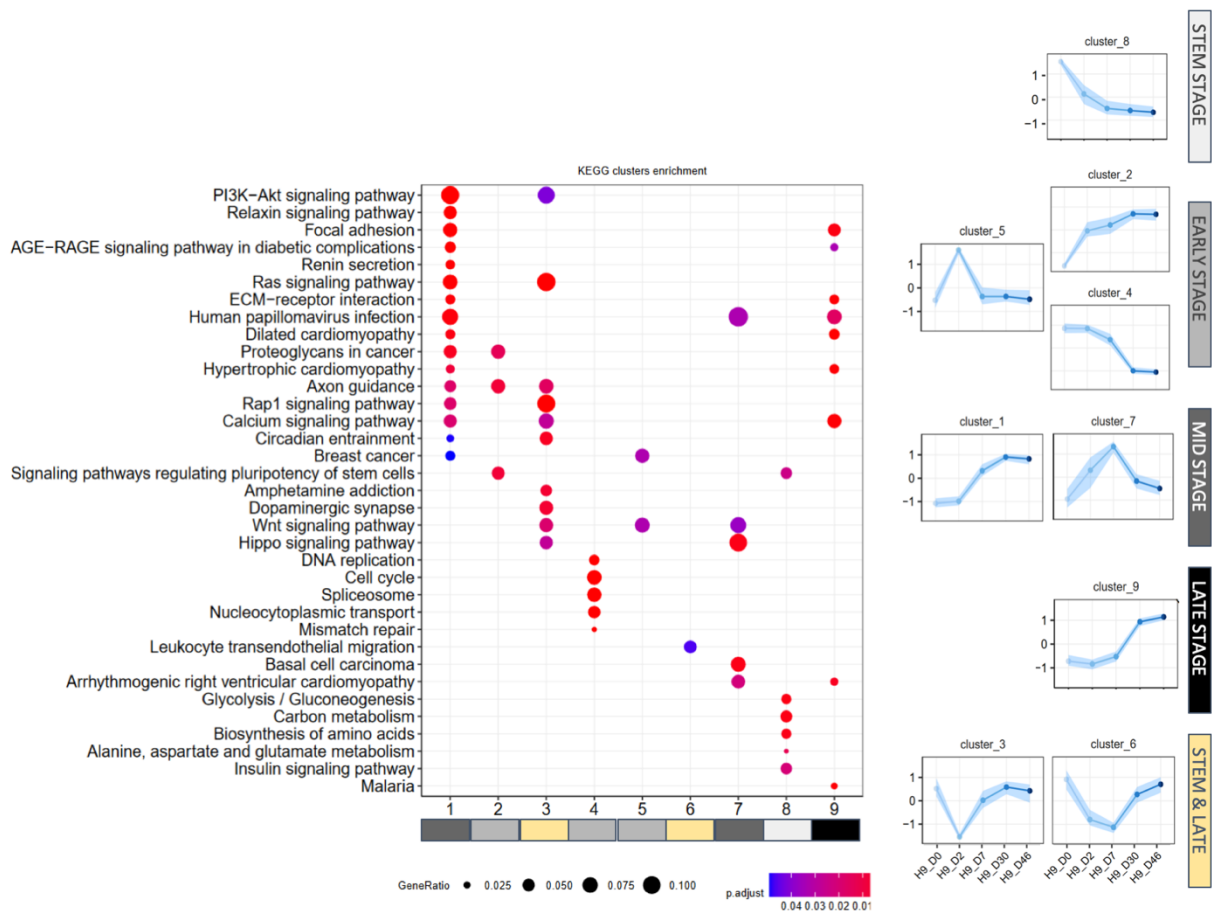


Figure 16. Cluster profiling based on KEGG pathways during H9 hESCs differentiation to muscle cells.

3.3 Involvement of genes with/out transcription factors (TFs) and epigenetic factors (Epifactors) functioning as proto-oncogenes and tumor suppressors across muscle differentiation stages

Differentially expressed gene analysis revealed the significant altered expression of a huge number of genes during muscle differentiation stages. These genes showed alteration in all stages, or, in a specific stage/stages. On the basis of the ANOVA method, genes were considered differentially expressed if their expression varied in at least one condition. Moreover, based on Pairwise contrasts method, genes were considered differentially expressed in one specific condition compared to a reference condition. It is worth to note that our analysis found involvement of 715 transcription factors, 278 epigenetic factors, and 915 lncRNAs (Figure 17)

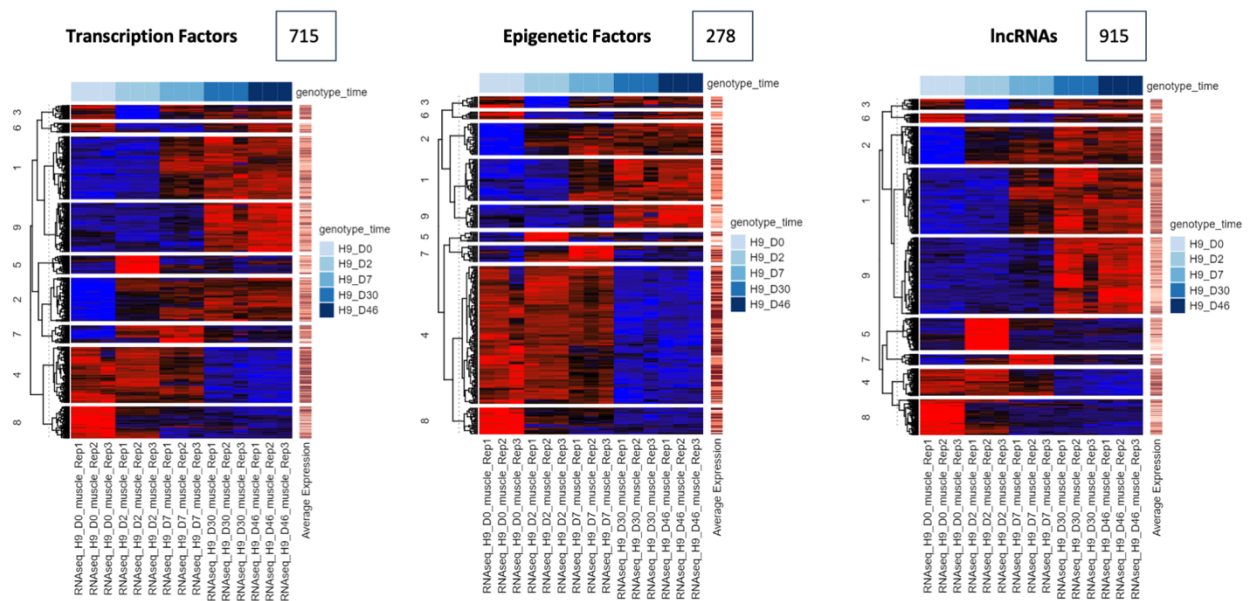


Figure 17. Heatmaps of differentially expressed genes as TFs, Epifactors, and also lncRNAs across muscle differentiation stages. It obviously shows different specific expression patterns of these factors across this process.

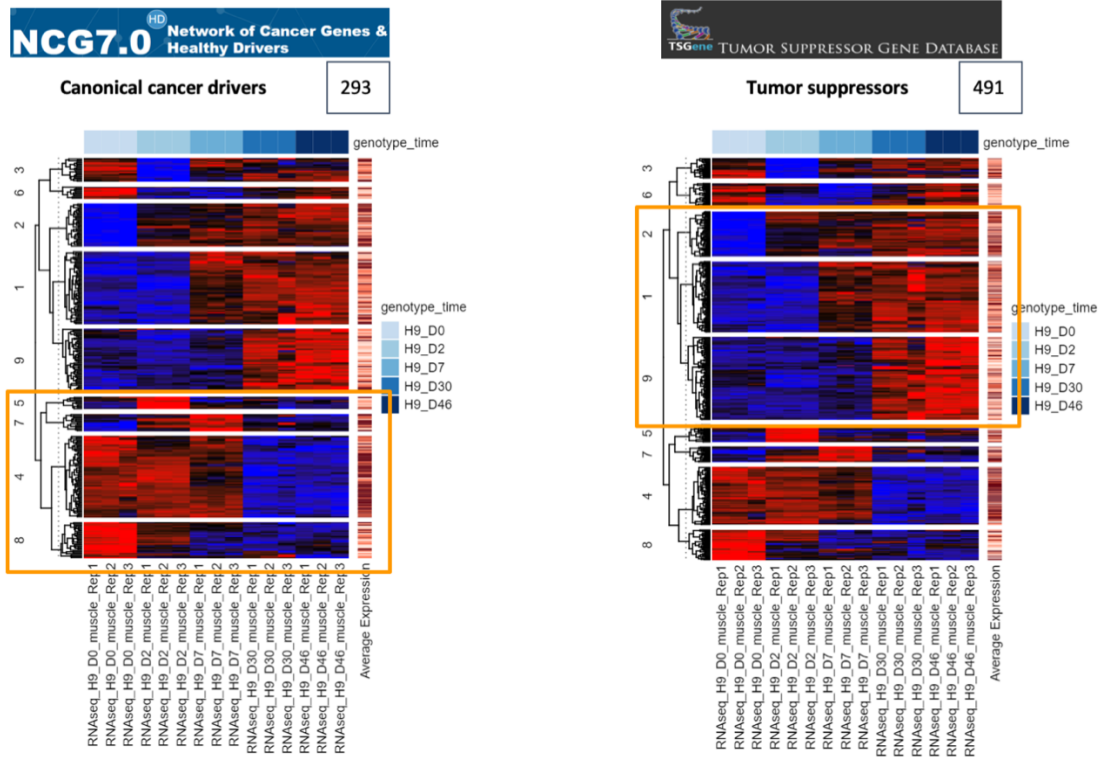
Regarding cancer drivers and tumor suppressors in our muscle differentiation stages, our results revealed 491 tumor suppressors based on TSGene database and 293 canonical cancer drivers based on NCG7 database (Figure 18A). The cluster profiling for these cancer-related genes showed that more cancer drivers were grouped in clusters 4, 5, 7, and 8 (Figure 18A) while tumor suppressors were among clusters 1,2, and 9 (Figure 18A).

3.3.1 Cancer drivers with proto-oncogenic and tumor suppressive roles from cluster 8

Regarding cancer drivers with proto-oncogenic and tumor suppressive roles in cluster 8, we found the significant alteration of several genes in all stages shown in Figure 18B and Table S1. Data revealed that all these important cancer drivers and tumor suppressors had lower expression in all stages of muscle differentiation compared to D0 (stemness). Their order is based on significant FDR. Among these genes, only *AXINI*, *PTPN6*, *FOXL2*, and *WNK2* have roles as both proto-oncogenes and tumor suppressors while other genes have role only as proto-oncogene (Figure 18B and Table S1). The most notable altered expression was for *POU5F1*, *USP44*, *LCK*, and among others. Genes with role as only TF were *POU5F* (as proto-oncogene), *ETV4* (as proto-oncogene), *MYC* (as proto-oncogene), *OLIG2* (proto-oncogene), *FOXL2* (proto-oncogene and tumor suppressor), *RGS7* (as proto-oncogene), and *BCL11A* (as proto-oncogene) (Table S1).

Among these genes, there is an important gene (*BCOR* as a proto-oncogene) which usually show altered expression in muscle related cancers such as sarcoma (as internal tandem duplication (ITD) in its last exon, resulting its overexpression, and as fusions with other genes. *BCOR* downregulation was more significant at stages of D7, 30, and 46.

A



B

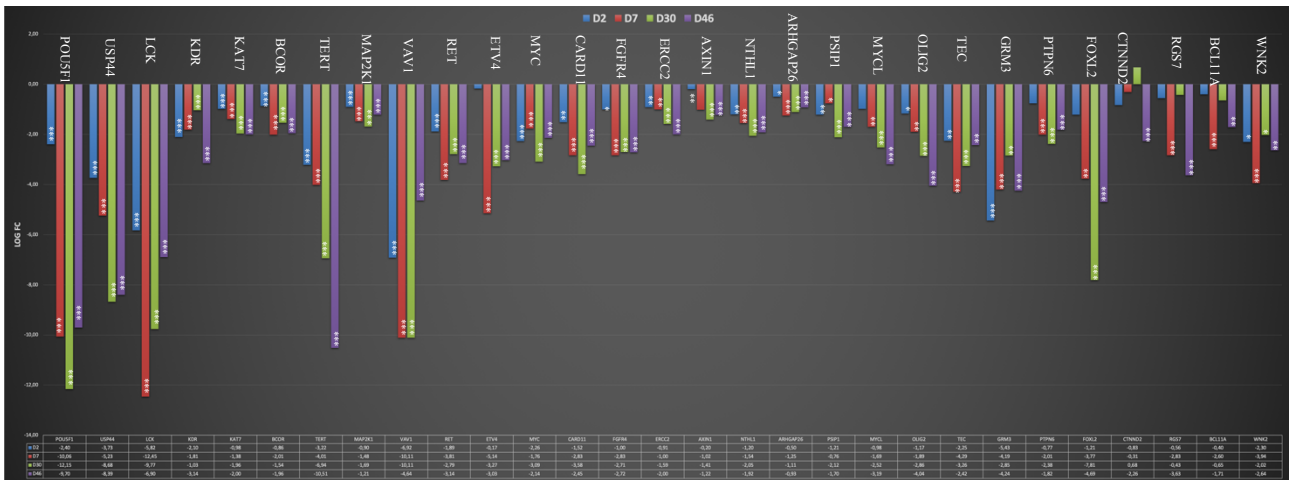


Figure 18. A. The cluster profiling for cancer-related genes across muscle differentiation stages. It represents key roles of tumor suppressor genes and canonical cancer drivers in muscle differentiation process. **B.** Cancer drivers with proto-oncogenic and tumor suppressive roles from cluster 8. The comparison between muscle differentiation stages and undifferentiated cells (D0) is based on log fold change. Genes are ordered based on their FDR significance. *: p-value less than 0.05; **: p-value less than 0.01; ***: p-value less than 0.001.

3.3.2 Cancer drivers with TF role as proto-oncogenes from clusters 4 and 8

Looking for cancer drivers in clusters 4 and 8, we identified involvement of several proto-oncogenes in all stages of muscle differentiation with role only as TF given in Figure 19 and Table S2 (the order is based on FDR significance). All these genes showed lower expression across almost all muscle differentiation stages compared to D0, representing their decreased expression in differentiated cells compared to stemness stages.

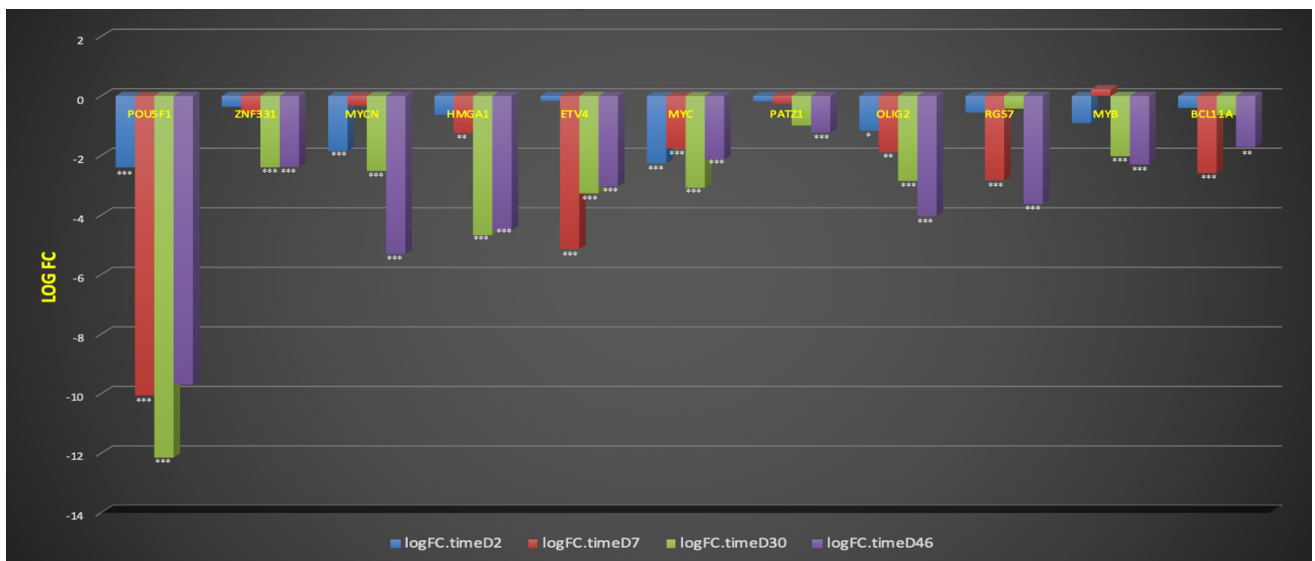


Figure 19. Involvement of several genes as proto-oncogenes in all stages of muscle differentiation with TF role. The comparison between muscle differentiation stages and undifferentiated cells (D0) is based on log fold change and their FDR significance. *: p-value less than 0.05; **: p-value less than 0.01; ***: p-value less than 0.001.

3.3.3 Involvement of cancer drivers with only TF role as proto-oncogenes across muscle differentiation stages

Looking for TF with only proto-oncogenic role, we found that several genes had altered expression in all stages compared to D0, grouped in clusters 4, 5, 7, and 8 (Table S3 and Figure 20), with more significant for *POU5F1* (lower expression) and *LEF1* (overexpression). The order of genes is based on the significance of FDR value. Among these genes, some important ones were *MYC* (with lower expression at all stages), *MYCN* (significant lower expression on D2, D30, and D46), *MYB* (significant lower expression on D30 and D46) and *PAX3* overexpression in all stages compared to D0.

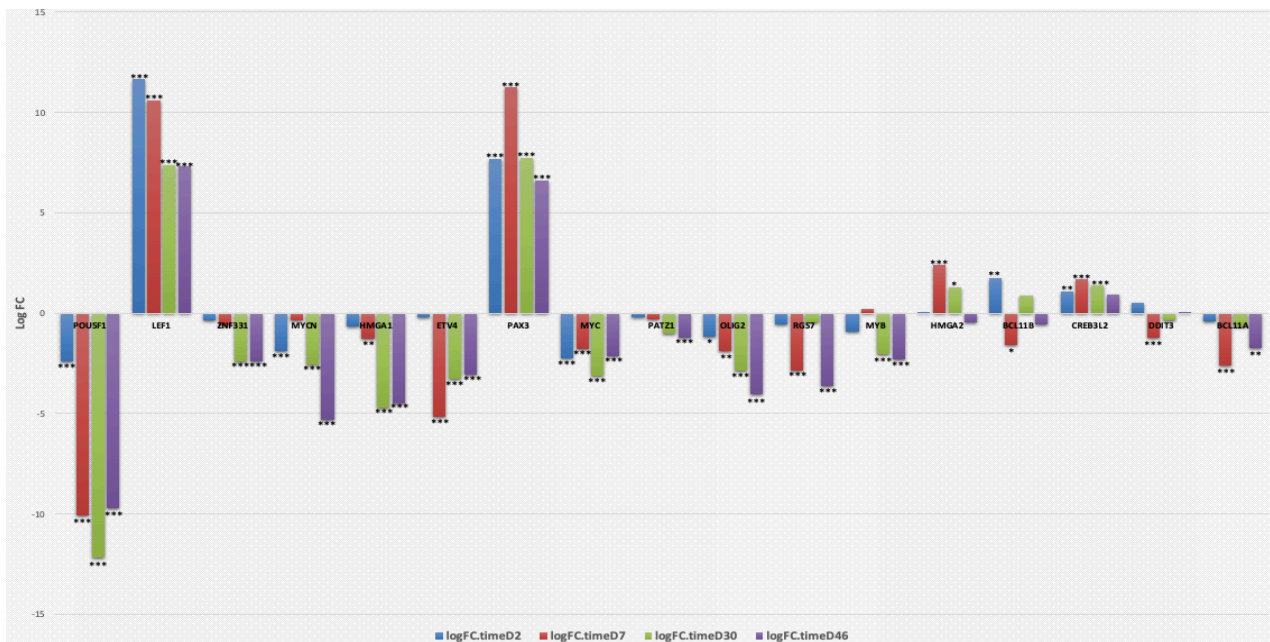


Figure 20. Involvement of TF genes as only proto-oncogenic role with altered expression in across muscle differentiation stages. The comparison between muscle differentiation stages and undifferentiated cells (D0) is based on log fold change and their FDR significance. *: p-value less than 0.05; **: p-value less than 0.01; ***: p-value less than 0.001.

3.3.4 Cancer drivers with Epifactor role with proto-oncogenic function across muscle differentiation stages

Genes with Epifactor role with proto-oncogenic function were mainly observed in clusters 4, 7, and 8. The most significant alteration was for *USP44* which showed significant down regulation across all stages. All genes in this group, except for *BAZ1A* (overexpression with more significant value on D 2 and D7, mesoderm and somite stages, respectively), were downregulated across almost all stages (Figure 21 and Table S4). Again, this data confirms the important roles of these proto-oncogenes in stemness status compared to the differentiated cells.

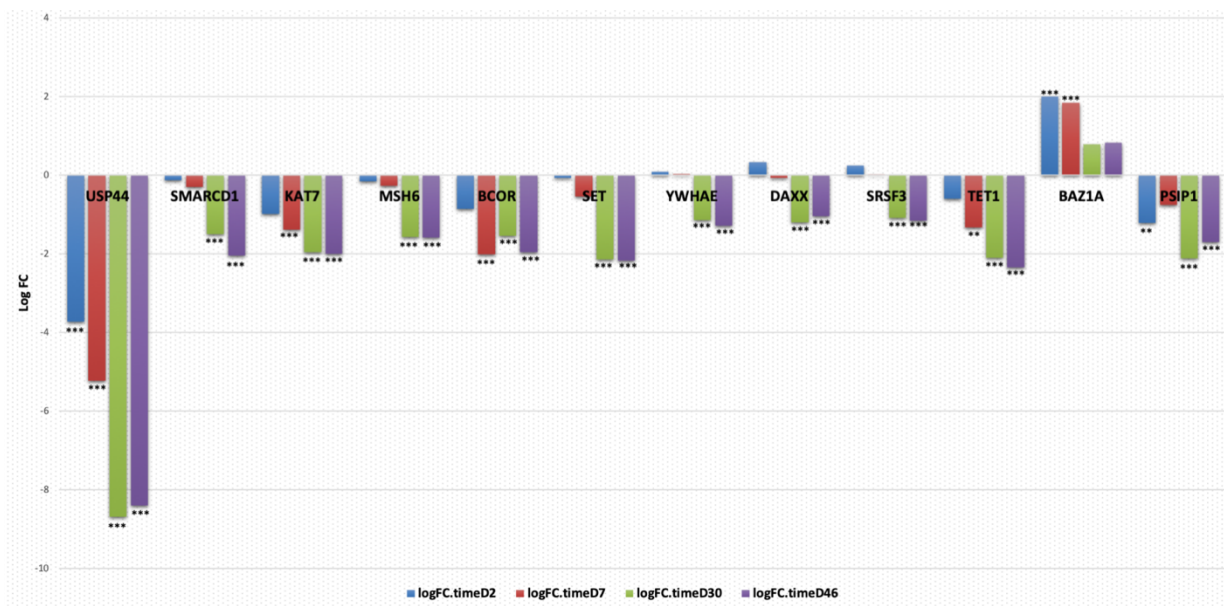


Figure 21. Involvement of genes with Epifactor role with proto-oncogenic function across muscle differentiation stages. The comparison between muscle differentiation stages and undifferentiated cells (D0) is based on log fold change and their FDR significance. *: p-value less than 0.05; **: p-value less than 0.01; ***: p-value less than 0.001.

3.3.5 Cancer drivers with both TF and Epifactor roles with both proto-oncogenic and tumor suppressive functions across muscle differentiation stages among clusters 4 and 7

In relation to the genes which have both TF and Epifactor roles and function as proto-oncogenes and tumor suppressors, we found four genes located in clusters 4 and 7. The most significant alteration was for *ZBTB16* with significant overexpression from D7 toward terminal stage. *EZH2* showed significant higher expression on D2 (mesoderm stage) and D7 (somite stage). *PBRM1* overexpression was more significant on D2 and *TP53* revealed more significant lower expressions on D30 and D46 (Figure 22).

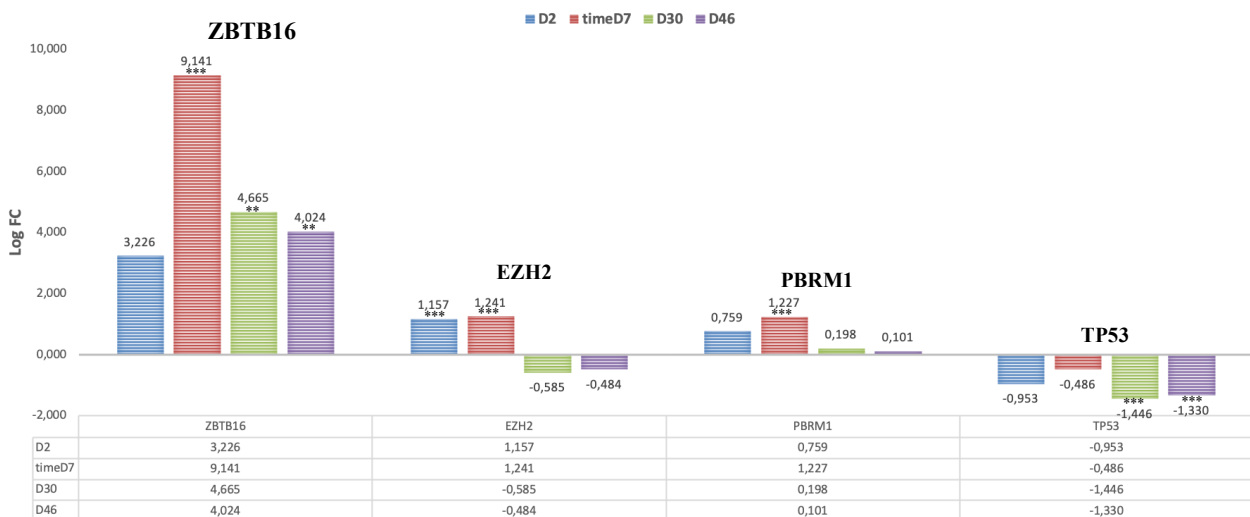


Figure 22. Altered expression of cancer drivers with both TF and Epifactor roles with proto-oncogenic function across muscle differentiation stages. The comparison between muscle differentiation stages and undifferentiated cells (D0) is based on log fold change and their FDR significance. *: p-value less than 0.05; **: p-value less than 0.01; ***: p-value less than 0.001.

3.3.6 Cancer driver genes with both tumor suppressive and proto-oncogenic roles which function as TF and/or Epifactor among clusters 4, 7, and 8

Among cancer drivers, we found significant altered expression of some genes which function as both tumor suppressors and proto-oncogenes (from clusters 4, 7, and 8) as follow (Figure 23): *SALL4* (more significant lower expression on D30 and 46), *CDX2* (more significant overexpression on D2 and D7), *CAMTA1* (significant lower expression on D46), *FOXL2* (significant lower expression from D7 toward terminal stage), and following genes as mentioned in previous section: *ZBTB16*, *EZH2*, *PBRM1*, and *TP53*.

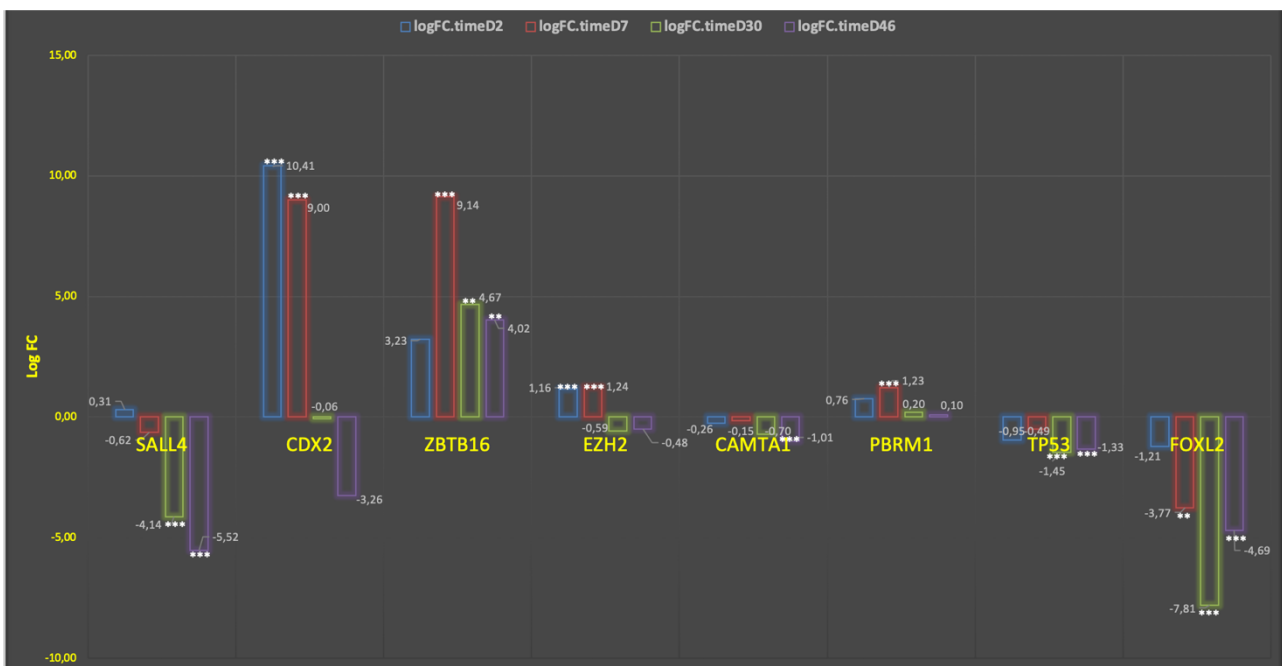


Figure 23. Altered expression of cancer driver genes with both tumor suppressive and proto-oncogenic roles functioning as TF and/or Epifactors. The comparison between muscle differentiation stages and undifferentiated cells (D0) is based on log fold change and their FDR significance. *: p-value less than 0.05; **: p-value less than 0.01; ***: p-value less than 0.001.

3.3.7 Cancer driver genes with only TF role which function as tumor suppressors

Cancer drivers extracted from clusters 4,5,7, and 8 revealed some TFs with tumor suppressive roles, with main alterations for overexpression of *GATA4* at all stages, lower expression of *SALL2* at all stages, and significant lower expression of *BCL6B* on D7, 30, 46 but its higher expression on D2 (Figure S1). Other genes are shown in this figure.

3.3.8 Cancer drivers with Epifactor role which function as tumor suppressors

Our data revealed that some cancer drivers from clusters 4,5, and 8 act as only tumor suppressors. These genes showed almost all lower expression across muscle differentiation (Figure 24). The main important gene was *DNMT3B* which showed very significant lower expression at all stages compared to stemness stage. While most of these genes showed down regulation in the most of stages, *ANP32A* and *PRKCD* revealed significant upregulation on D2 (mesoderm stage) (Figure 24).

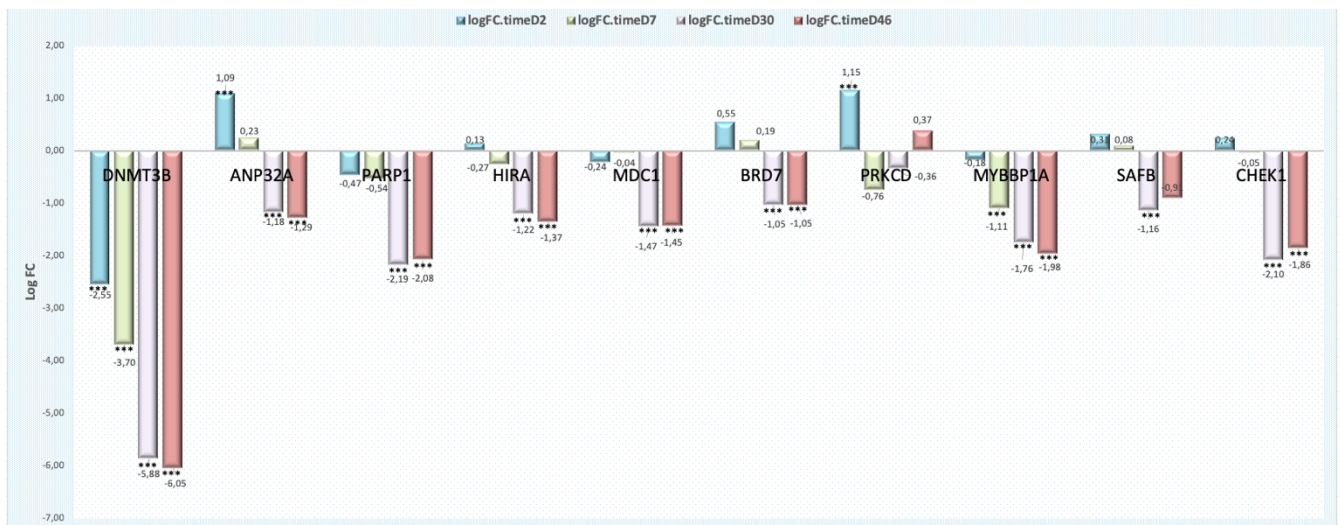


Figure 24. Involvement of cancer drivers with Epifactor role functioning as only tumor suppressor. The comparison between muscle differentiation stages and undifferentiated cells (D0) is based on log fold change and their FDR significance. *: p-value less than 0.05; **: p-value less than 0.01; ***: p-value less than 0.001.

3.3.9 Cancer drivers with roles of both TF and Epifactor which function as tumor suppressors from cluster 4

We found down regulation of two cancer drivers with both TF and Epifactor roles functioning as tumor suppressors, including *CTCF* (significant lower expression at all stages with following logFC: D2 vs D0: -1.81; D7 vs D0: -2.39; D30 vs D0:-4.99; and D46 vs D0: -9.10) and *SMARCC1* (significant lower expression at D30, logFC:-1.16; and D46, logFC:-1.32) (Figure S2). There were enriched in the cluster 4.

3.3.10 Tumor suppressor genes with TF role

As seen in figure S3, almost all of these TF genes which function as tumor suppressors (grouped in clusters 1, 2, and 9) showed higher expression across muscle differentiation stages, with the most significant higher expressions for *IFI16* and *FOXCI* across all stages compared to stemness stage (D0) (Figure S3).

3.3.11 Tumor suppressor genes with Epifactor role

Our results found several important genes with altered expression across muscle stages which were in clusters 1, 2, and 9. The main significant genes were *SMARCA2* with higher expression at all stages, and *EYAA4* and *SIRT2* with significant higher expression from D7 toward late stage compared to stemness stage (D0) (Figure 25).

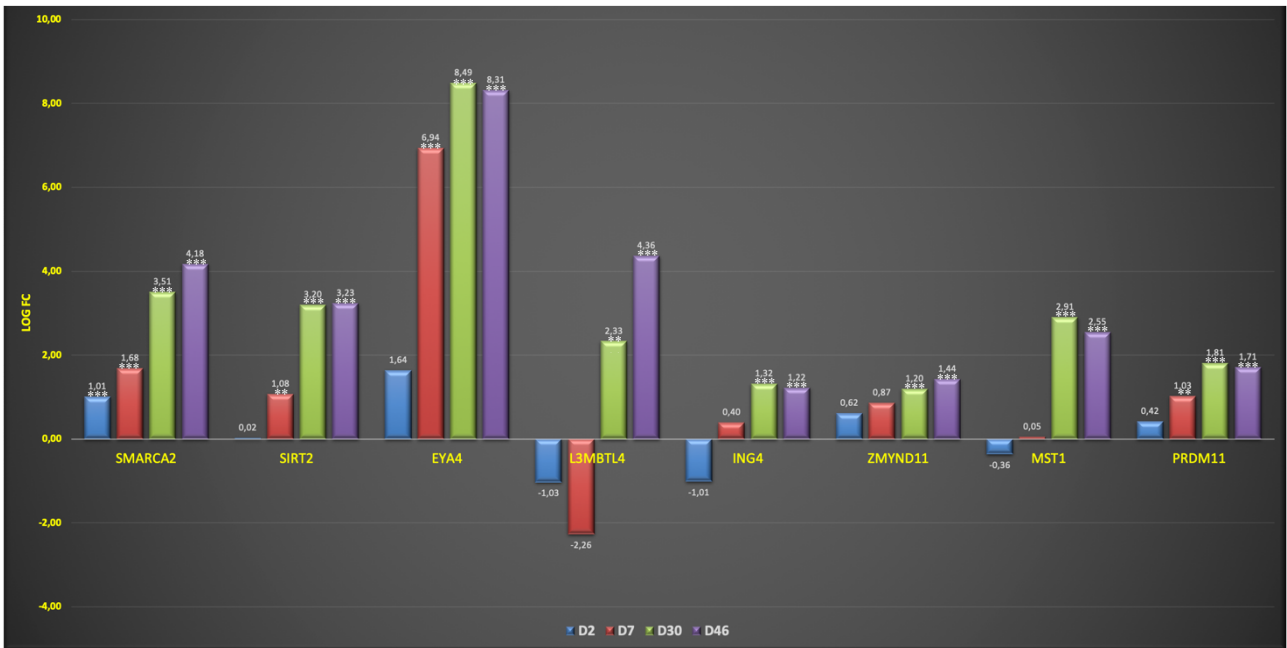


Figure 25. Altered expression of tumor suppressor genes with Epifactor role across muscle differentiation. The comparison between muscle differentiation stages and undifferentiated cells (D0) is based on log fold change and their FDR significance. *: p-value less than 0.05; **: p-value less than 0.01; ***: p-value less than 0.001.

3.3.12 Tumor suppressor genes with both Epifactor and TF roles

Regarding genes with role of both TF and Epifactors functioning as tumor suppressors grouped in cluster 1 and 9, we found three genes with altered expression across muscle differentiation stages, with more significant overexpression for *SP100* from D7 toward terminal stage compared to D0. *NPAS2* revealed significant down regulation on D2 but significant upregulation on D30 and D46. *JDP2* only showed significant overexpression on D46 (late stage) (Figure26).



Figure 26. Involvement of tumor suppressor genes with both Epifactor and TF roles across muscle differentiation. The comparison between muscle differentiation stages and undifferentiated cells (D0) is based on log fold change and their FDR significance. *: p-value less than 0.05; **: p-value less than 0.01; ***: p-value less than 0.001.

3.3.13 Genes with roles of both TF and Epifactor which function as both proto-oncogenes and tumor suppressors across all clusters

Our data revealed that during muscle differentiation stages, there were altered expression of genes (in clusters of 1, 3, 4, and 7) classified as both TFs and Epifactors which have both proto-oncogenic and tumor suppressive functions. The more significant alterations were for *FOXP1* (overexpression on D7, 30, and 46), *FOXO1* (downregulation at all stages), and *ZBTB16* (as mentioned before, overexpression from D7 toward late stage), in order (Figure 27).

Moreover, as mentioned before, *EZH2* showed significant overexpression on D2 and D7 but *PBRM1* overexpression was on D7. *TP53* revealed significant down regulation at late stages (D30 and 46) and *PRDM1* showed over expression from D7 upward (Figure 27).

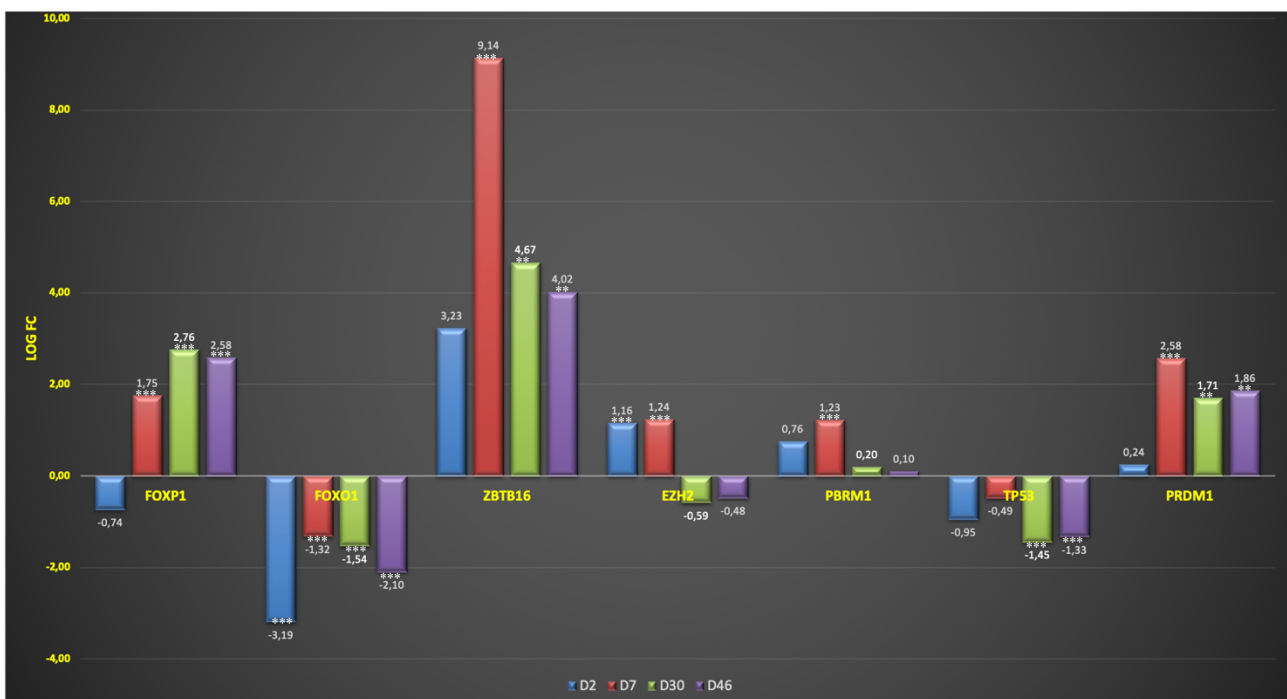


Figure 27. Involvement of genes with role of both TF and Epifactor functioning as both proto-oncogenes and tumor suppressors across. The comparison between muscle differentiation stages and undifferentiated cells (D0) is based on log fold change and their FDR significance. *: p-value less than 0.05; **: p-value less than 0.01; ***: p-value less than 0.001.

3.3.14 Proto-oncogenes without TF and Epifactor roles

In this category, our data revealed a number of genes from clusters of 4, 5, 7, and 8 which altered across differentiation stages (Table S5), with more downregulation at late stages as expected for proto-oncogenes. However, more significant changes were found for downregulation of *LCK*, *BUB1B*, *RMI2*, and *KDR* at all stages and *RHOH* on D7, 30, and 46; and overexpression of *COL2A1* at all stages and *RHOH* on D2 (Figure 28 and Table S5).

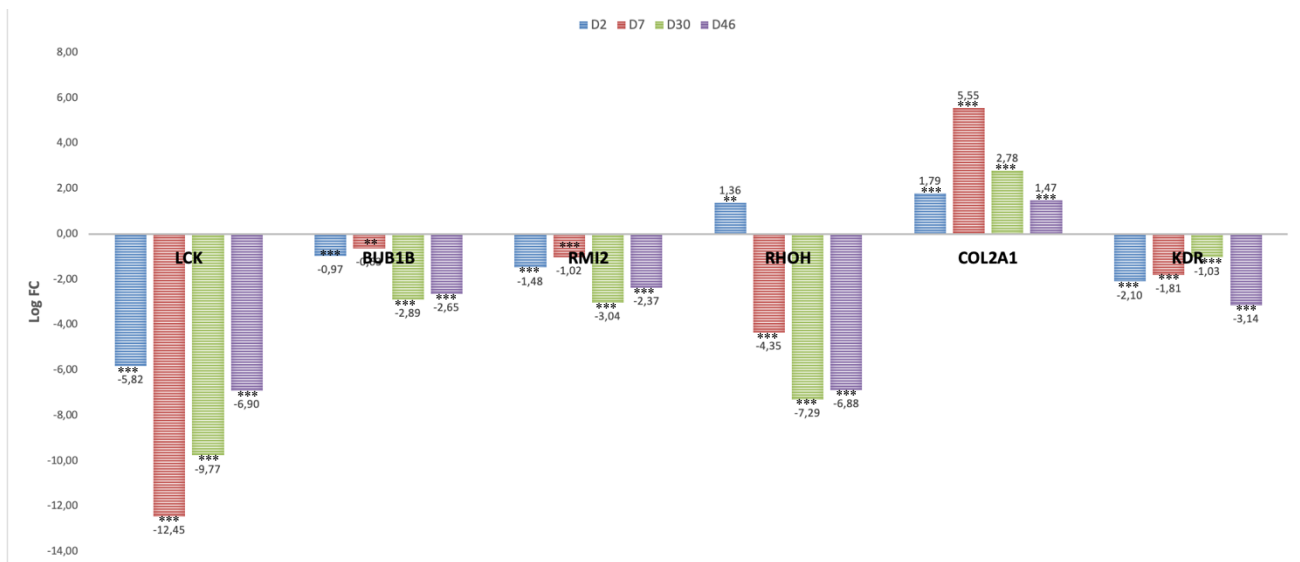


Figure 28. Significant altered expression of some genes functioning as proto-oncogenes. The comparison between muscle differentiation stages and undifferentiated cells (D0) is based on log fold change and their FDR significance. *: p-value less than 0.05; **: p-value less than 0.01; ***: p-value less than 0.001.

3.3.15 Tumor suppressor genes without TF and Epifactor roles

We also found a number of genes (174) without TF and Epifactor roles which function as tumor suppressors from clusters of 1, 9, and 2 (Table S6). Most of these genes revealed overexpression across muscle differentiation as expected for tumor suppressors. Some of the most significant genes are also given in Figure 32. The most significant alteration was for *EMP1* with downregulation on D2 but upregulation on D7, 30, and D46. Almost all genes showed overexpression across differentiation stages for example, *PCDH9*, *VIM*, *WNT5A*, and *AHNAK* revealed significant overexpression at all stages while *GAS1*, *DMD*, and *IGFBP5* overexpression on D7, 30, and 46. However, *IGFBP5* showed significant lower expression on D2. Moreover, *SRPX* revealed significant upregulation on D30 and 46 while significant *NUPR1* overexpression was on D2, 30, and 46 (Figure 29). The altered expression of other genes is given in Table S6.

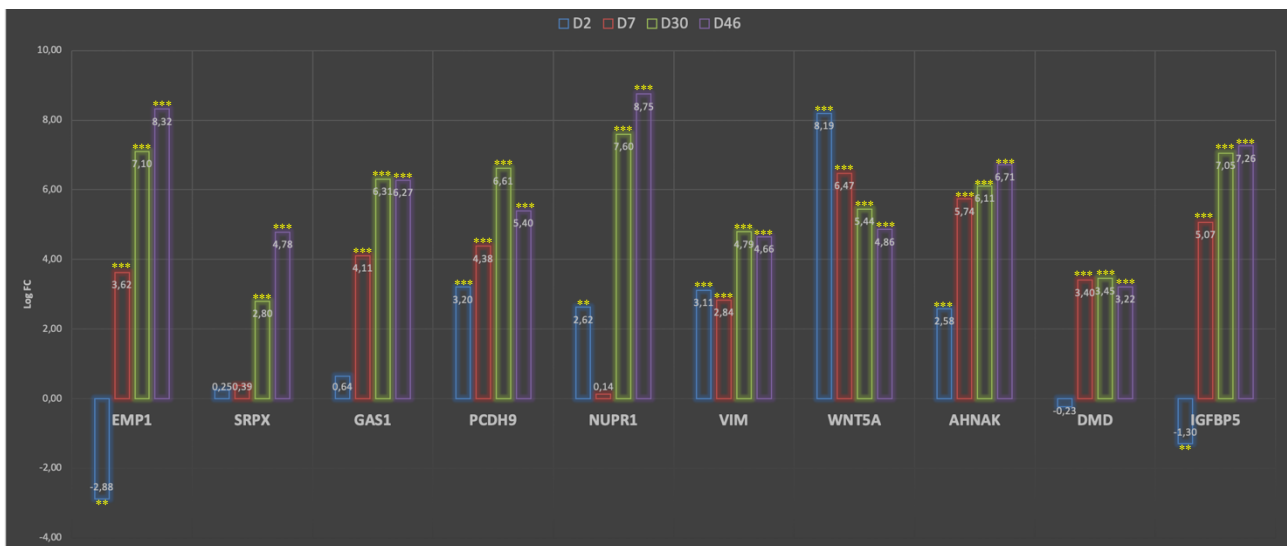


Figure 29. Significant altered expression of some tumor suppressors without roles of TF and Epifactor. The comparison between muscle differentiation stages and undifferentiated cells (D0) is based on log fold change and their FDR significance. *: p-value less than 0.05; **: p-value less than 0.01; ***: p-value less than 0.001.

3.3.16 Tumor suppressor genes and proto-oncogenes without TF and Epifactor roles among all clusters

Our results revealed several main genes with function of both tumor suppressive and proto-oncogenic roles but not as TF and Epifactors. Most of them showed over expression during muscle differentiation stages. The most significant alteration was for *EPHA3* with down regulation on D2 but upregulation at other stages (D7, 30, and 46) compared to D0. Some of other important genes which are also involved in cancers included *TGFBR2*, *CDKN2C*, *DCC*, *BTK*, *FAS*, *NOTCH2*, *CHEK2*, *BLM*, *CDKN1A*, *MSH2*, *CDH1*, *PTCH1*, *CYLD*, *CASP8*, *NOTCH1*, *CBL*, *PML*, *CDKN2A* (Figure S4 and Table S7), and among others. Among these genes in this category, some of them showed alteration at all stages as follow: *CDH11*, *DCC*, *LRIG3*, *DDR2*, *CDK6*, and *CDKN2A* with overexpression at all stages; *CDH1* and *WNK2* lower expression at all stages, *EPHA3* with significant downregulation on D2 and over expression on D7, 30, and 46; *BTK* with significant overexpression on D2 but lower expression on D7, 30, 46; and *WIF1* with significant down regulation on D2 and D7 but upregulation on D30 and D46. All other genes with altered expression are given in Table S7.

3.3.17 The involvement of genes with role as both TF and Epifactor in muscle differentiation stages

In this part of expression analysis, we looked for the expression of genes with both TF and Epifactor roles across muscle differentiation stages (Figure S5A). We found several genes with altered expression at all stages (Figure S5 A and B), specific changed expression at a specific stage, two or more stages (Figure S5 A and 30A-D), helping identify altered pathways in these stages. Our results revealed that some of these genes had altered expression (down ↓ or upregulation ↑) at all muscle differentiation stages compared to D0 as follow (Figure S5 A and B): *CTCF* (↓), *FOXO1* (↓), *PRDM14* (↓), *DPF3* (↑), *PRDM8* (↑), *SCML4* (↑), *SNAI2* (↑), *ZHX1* (↑), and *ZNF516* (↑). Based on Enrichr database (<https://maayanlab.cloud/Enrichr/enrich#>), the biological processes (GO:BP) related to these genes with most important one was regulation of transcription, DNA-templated (GO:0006355) (Figure 31A and Table S8). The main enriched genes for these pathways were *SNAI2*, *FOXO1*, and *ZNF516*, in order (Figure 31A).

In this category (genes with both TF and Epifactor roles), the genes with only altered expression on D2 (paraxial mesoderm) were as follow (Figure S5A and 30A): *PCGF6* (↑), *SATB1* (↓), and *ZNF217* (↓). Based on Enrichr database, more enriched genes were *PCGF6*, *ZNF217*, and *SATB1* (Figure 31B). The main enriched GO:BPs included negative regulation of transcription, DNA-templated (GO:0045892), regulation of transcription by RNA polymerase II (GO:0006357), regulation of transcription, and DNA-templated (GO:0006355); then for negative regulation of G0 to G1 transition (GO:0070317) and regulation of G0 to G1 transition (GO:0070316) (Figure 31B and Table S9).

Regarding changed expression of the genes with both TF and Epifactor roles at somite stage (D7), *ATF2* (↑), *CHD1* (↑), *PBRM1* (↑), *SMARCA1* (↑), and *HMG20B* (↓), showed only altered expression at this stage (Figure S5A and 30B). Enrichr database revealed that the more enriched genes were *CHD1*, *ATF2*, and *PBRM1* in order. *ATF2* is involved in positive regulation of myoblast proliferation (Figure 31C and Table S10).

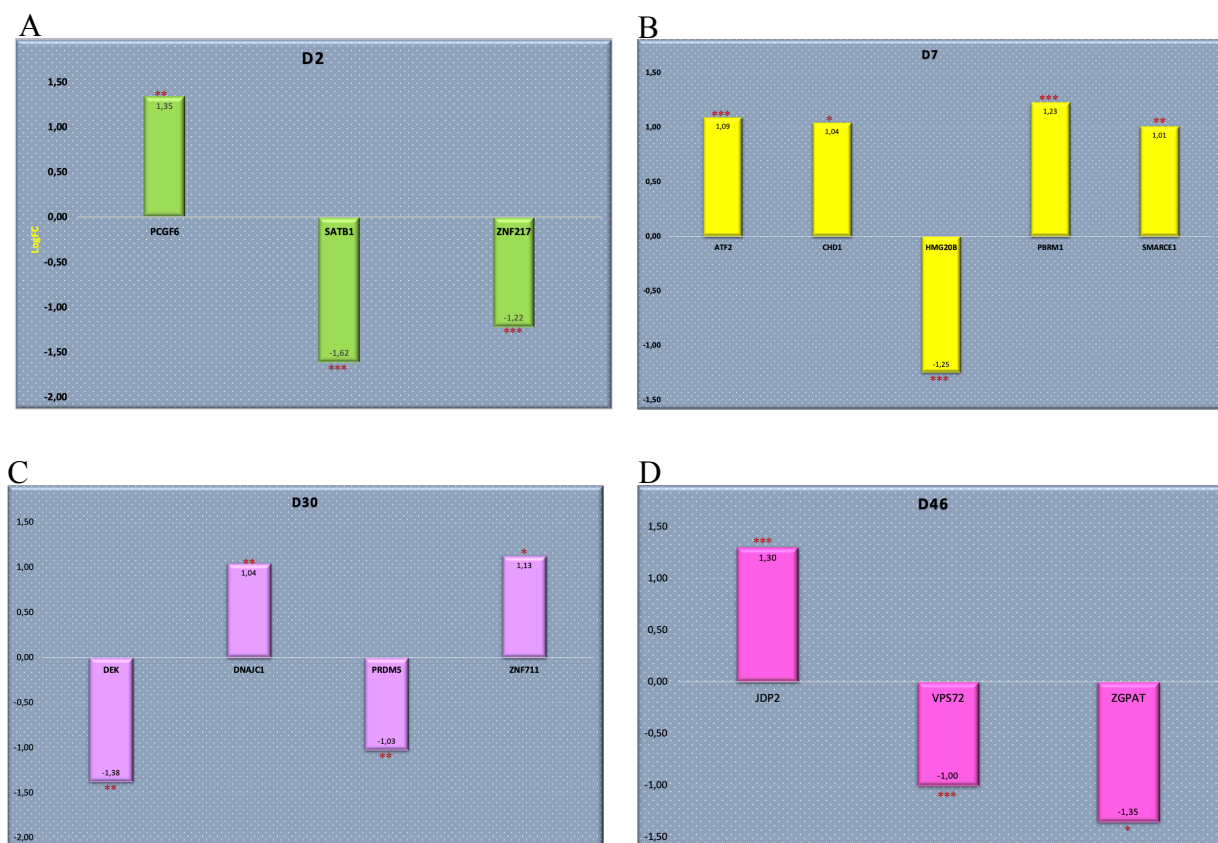


Figure 30. Genes with both TF and Epifactor functions changed only at specific stage (A: D2, B: D7, C: D30, and D: D46) of muscle differentiation process compared to undifferentiated cells (D0). The comparison between muscle differentiation stages and undifferentiated cells (D0) is based on logFC. *: p-value less than 0.05; **: p-value less than 0.01; ***: p-value less than 0.001.

Some of these genes only had altered expression on D30 which included *DEK* (↓), *PRDM5* (↓), *ZNF711* (↑), and *DNAJC1* (↑) (Figure S5A and 30C). Enrichr database showed that the main enriched genes were *PRDM5* (involved in histone H3-K9 methylation (GO:0051567), histone lysine methylation (GO:0034968), regulation of macromolecule metabolic process (GO:0060255), protein deacetylation (GO:0006476), and histone deacetylation (GO:0016575)) and *DEK* (Figure 31D and Table S11).

In relation to the terminal differentiation (D46), *JDP2* (↑), *VPS72* (↓), *ZGPAT* (↓) showed only alteration at this stage (Figure S5A and 30D). Enrichr GO:BP identified *ZGPAT* as the major enriched gene involved in regulation of transcription, epidermal growth factor-activated receptor activity, protein tyrosine kinase activity, and epidermal growth factor receptor signaling pathway (Figure 31E and Table S12).

Other combinations of altered expression can be observed in Figure S5A. For example, there were genes with similar altered expression, down- or up-regulation, only at both first stages (D2 and D7) such as *EZH2* (↑) and *VDR* (↓). Enrichr database revealed that both these genes are involved in several main GO: BP terms such as regulation of gene expression, multicellular organismal process, nucleic acid-templated transcription, cellular macromolecule biosynthetic process, cellular response to vitamin D, retinoic acid receptor signaling pathway, calcidiol 1-monooxygenase activity, keratinocyte proliferation, chromatin silencing at telomere, and regulation of transcription by RNA polymerase II (Table S13).

Genes with similar altered expression only at both late stages (D30 and D46) included *ASHIL* (↑), *BRPF1* (↓), *CBX2* (↓), *CHD7* (↓), *CHD9* (↑), *DNAJC2* (↓), *DOTIL* (↓), *GATAD2A* (↓), *MBD2* (↑), *PRDM16* (↑), *RAD51* (↓), *RAG1* (↑), *RARA* (↑), *SMARCA1* (↑), *SMARCA5* (↓), *SMARCC1* (↓), *SSRP1* (↓), *TP53* (↓), *YEATS4* (↓), *ZBTB7C* (↑), and *ZNF687* (↓). Based on the Enrichr database, we found main enriched GO:BP terms of these genes for regulation of nucleic acid-templated

transcription, ATP-dependent chromatin remodeling, transcription by RNA polymerase II, chromatin organization, among other (Figure S6A and Table S14).

One of several interesting findings was about altered expression of forkhead-box (*FOX*) gene family, *FOXA1*, *FOXP1*, *FOXP2*, and *FOXP4*; and some other genes such as *HP1BP3*, *NCOA1*, *PRDM1*, *PRDM6*, and *RCOR3* across muscle differentiation. They revealed only overexpression from D7 toward terminal differentiation (D30 and D46) compared to D0, representing their important roles in this differentiation process. Again based on the Enrichr database, the most significant enriched GO:BP terms for this category were regulation of transcription by RNA polymerase II (GO:0006357) and regulation of transcription, DNA-templated (GO:0006355) with involvement of all above-mentioned *FOX* genes enriched in the two terms. Overall, the main enriched genes were *FOXP1* and *FOXP2* (Figure S6B and Table S15).

In this category, the gene that showed altered expression (overexpression) only on D2, 7 and D30 was *SATB2* with involvement in chromatin remodeling (GO:0006338) and chromatin organization (GO:0006325) extracted from the Enrichr database.

3.3.18 The involvement of genes with role as only Epifactor in muscle differentiation stages

In another investigation in differential expression of muscle differentiation stages, we looked for genes with only function as Epifactor. Regarding the role of main Epifactors, DNA methyltransferases (DNMTs), our results revealed the down regulation of *DNMT3B* at all stages compared to D0 (stem cell) while *DNMT3A* only downregulated on D2 (mesoderm stage) and *DNMT1* only on D30 and 46 (late myogenic stages) (Figure 32). This data represents the major role of *DNMT3B* in stemness status compared to differentiated cells. Moreover, it shows that *DNMT3A* expression may important to compensate the expression of other DNA methyltransferases at different muscle differentiation stages.

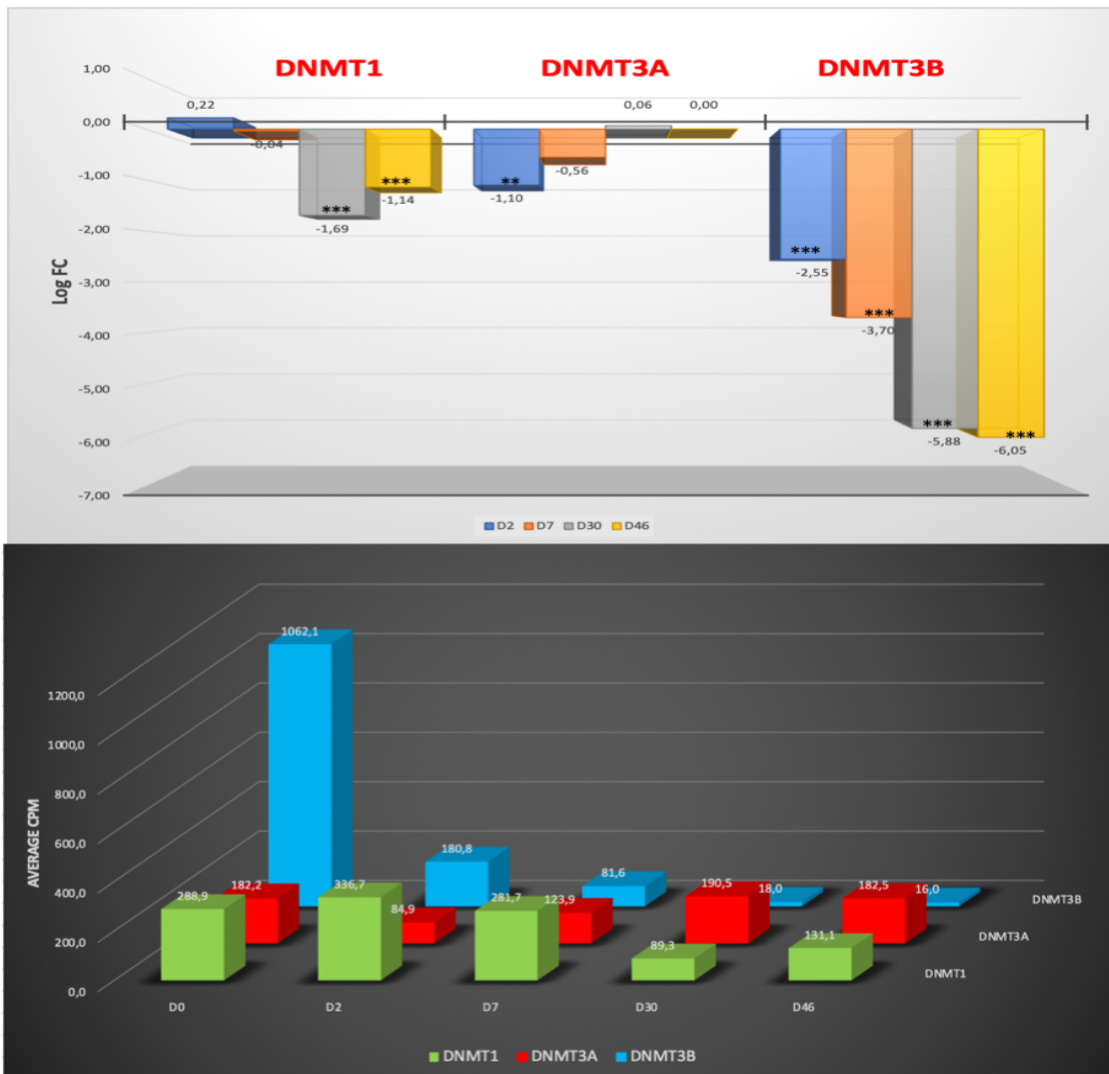


Figure 32. Altered expression of DNMTs across muscle differentiation stages. The comparison between muscle differentiation stages and undifferentiated cells (D0) is based on logFC and their FDR significance for upper part and on average CPM for lower part. *: p-value less than 0.05; **: p-value less than 0.01; ***: p-value less than 0.001.

Overall, in relation to the genes with only Epifactor role, some genes showed significant altered expression at all stages of muscle differentiation as follow: **Gene with higher expression at all stages compared to D0:** *ASXL1, BMI1, CBX8, CHD3, DND1, ERBB4, FBRS, JAK2, MLLT6, MSL3, NCOA6, PCGF5, PHC2, PHF20L1, RAI1, RBI, SAP30L, SETD7, SMARCA2, SMARCD3, SMYD2, TET2, YAF2, and ZCWPW1*. **Gene with lower expression at all stages compared to D0:** *CBX7, DNMT3B, JADE1, JARID2, NAP1L2, PADI3, PHC1, PKN1, TDRKH, and USP44* (Figure S7 and S8). Enrichr GO:BP terms analysis revealed that in this category, the main enriched GO:BP were negative regulation of transcription, DNA-templated and regulation of transcription by RNA polymerase II with enrichment of most of these genes in these two terms. The major enriched genes were *SMARCA2, RBI, and CHD3*, in order (Figure 33A and Table S16).

It is worth to note that among the genes in this category with only altered expression at all stages, only *DTX3L* showed lower expression on D2 and D7 (mesoderm and somite stages) but higher expression on D30 and D46 (myogenic progenitors and terminal stages) (Figure S7 and S8). This gene is involved in regulation of receptor catabolic process, protein localization to early endosome and endosome, chromatin binding, histone H2B ubiquitination, ubiquitin-protein transferase activity, double-strand break repair via nonhomologous end joining, and among others (Table S17, extracted from Enrichr database).

In addition, some genes as only Epifactor role showed only altered expression on D2 (paraxial mesoderm stage) which included *CHUK, HLCS, MEAF6, PRMT8, and WSB2* with higher expression; and *DNMT3A, HDAC7, HDAC9, HLTF, INO80D, PRKAA2, PRKCA, TLE1, and UBE2H* with lower expression (Figure S7 and Figure 34A). Regarding this category, the main enriched GO:BP term from Enrichr database was protein deacetylation (GO:0006476) and the most enriched gene was *HDAC9* involved in regulation of blood vessel endothelial cell migration, histone modification, cell migration involved in sprouting angiogenesis, intracellular signal transduction, and cellular protein modification process (Figure 33B and Table S18).

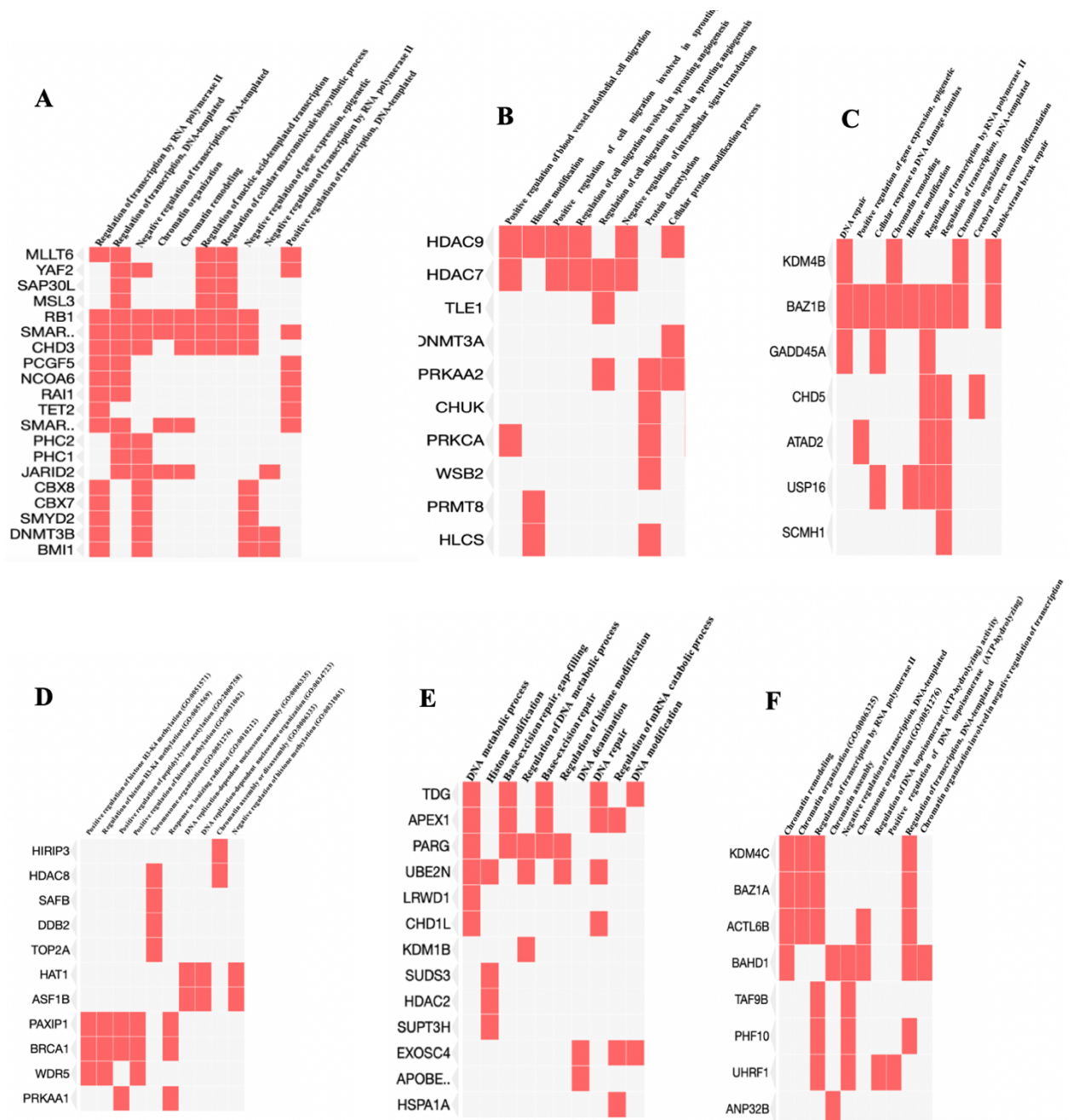


Figure 33. The enriched GO:BP for genes as only Epifactor. **(A)** genes with altered expression at all stages. **(B)** genes with only expression alteration on D2. **(C)** genes with only higher expression at somite stage (D7). **(D)** genes with only expression changes on D30. **(E)** genes with only changed expression at terminal stage (D46). **(F)** genes with altered expression only on both D2 and D7. GO:BP terms are listed based on adjusted p-value (the value less than 0.05 were considered statistically significant.)

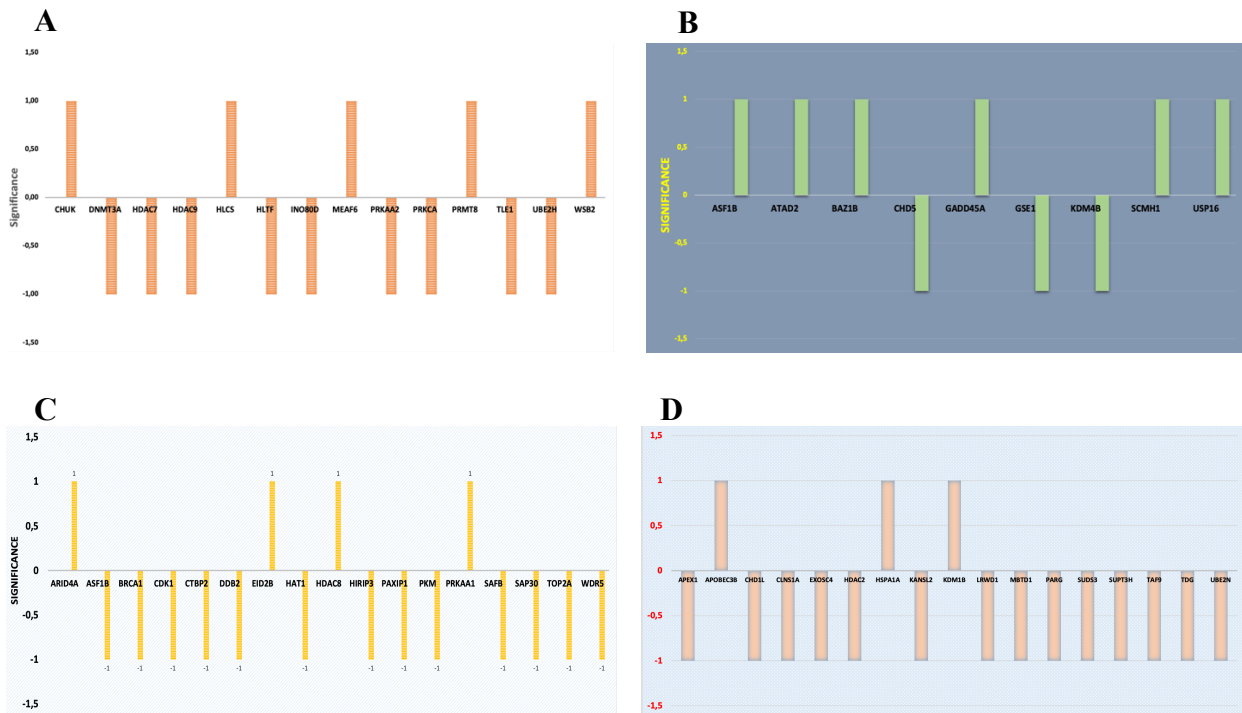


Figure 34. Altered expression of genes with only Epifactor role at all stages of muscle differentiation. The comparison between muscle differentiation stages and undifferentiated cells (D0) is based on significance of expression (+1 for overexpression and -1 for lower expression). **(A), (B), (C), and (D):** Genes specific for D2, D7, D30, and D46, respectively.

Genes as only Epifactor with only higher expression at somite stage (D7) included *ASF1B*, *ATAD2*, *BAZ1B*, *GADD45A*, *SCMH1*, and *USP16* while genes with only lower expression at this stage were *CHD5*, *GSE1*, and *KDM4B* (Figure S7 and Figure 34B). The main enriched gene in this category on the basis of Enrichr database was *BAZ1B* involved in DNA repair, regulation of gene expression, epigenetic, cellular response to DNA damage stimulus, chromatin remodeling, histone modification, regulation of transcription by RNA polymerase II (among the most enriched GO:BP), regulation of transcription, DNA-templated (among the most enriched GO:BP), chromatin organization, and double-strand break repair (Figure 33C and Table S19).

Among genes as only Epifactor with only changed expression on D30 were as follow: *ARID4A*, *EID2B*, *HDAC8*, *PRKAA1* (higher expression) and *ASF1B*, *BRCA1*, *CDK1*, *CTBP2*, *DDB2*, *HAT1*, *HIRIP3*, *PAXIP1*, *PKM*, *SAFB*, *SAP30*, *TOP2A*, and *WDR5* (lower expression) (Figure S7 and 34C). The main enriched Enrichr GO:BP terms were positive regulation of histone H3-K4 methylation (GO:0051571), regulation of histone H3-K4 methylation (GO:0051569), and positive regulation of histone methylation (GO:0031062) with involvement of *PAXIP1*, *BRCA1*, and *WDR5*; positive regulation of peptidyl-lysine acetylation (GO:2000758, with involvement of *PAXIP1*, *BRCA1*, and *PRKAA1*), and chromosome organization (GO:0051276, with involvement of *HDAC8*, *SAFB*, *DDB2*, *TOP2A*) (Figure 33D and Table S20).

At terminal stage (D46), we observed some important Epifactor genes with only changed expression at this stage as follow (Figure S7 and Figure 34D): *APOBEC3B*, *HSPA1A*, and *KDM1B* with upregulation; and *APEX1*, *CHD1L*, *CLNS1A*, *EXOSC4*, *HDAC2*, *KANSL2*, *LRWD1*, *MBTD1*, *PARG*, *SUDS3*, *SUPT3H*, *TAF9*, *TDG*, and *UBE2N* with downregulation. Their most enriched GO:BP terms were DNA metabolic process, histone modification, and DNA repair and the main enriched genes found from Enrichr database were *TDG*, *APEX1*, *PARG*, and *UBE2N* (these genes are involved in DNA metabolic process.) (Figure 33E and Table S21).

There were some Epifactor genes that showed altered expression only on D2 and D7 as follow (Figure S7): *ACTL6B*, *APOBEC3C*, *BAHDI*, and *KDM4C* with lower expression; and *ANP32B*, *BAZIA*, *PBK*, *PHF10*, *TAF9B*, and *UHRF1* with higher expression. The most enriched GO:BP terms were regulation of transcription by RNA polymerase II (GO:0006357), regulation of transcription, DNA-templated (GO:0006355), and chromatin remodeling (GO:0006338). The most enriched gene in this category was *BAHDI* involved in chromatin remodeling, assembly, and organization and regulation of transcription (Figure 33F and Table S22).

We found that several Epifactor genes showed expression changes from D7 toward the late stages (D30 and D46) with significant downregulation and upregulation (Figure S7 and Table S23). Most of them are enriched in GO:BP terms of regulation of transcription (Figure S9A). Another interesting

part was the overexpression or lower expression of lots of Epifactor genes only on both D30 and D46 (Figure S7 and Table S24). Most of them were enriched in GO:BP terms of chromatin remodeling and organization, nucleosome organization, cellular response to UV, regulation of signal transduction by TP53, DNA metabolic process, ribosome biogenesis, histone H4 acetylation, and nuclear polyadenylation-dependent rRNA catabolic process (Figure S9B).

3.4. Involvement of myokines in muscle differentiation

In this part of our study, we investigated the possible involvement of some important myokines at muscle differentiation stages, helping identify their uncharacterized biological pathways. Overall based on the late stage of differentiation, the most significant alterations were for overexpression of *DCN*, *MSTN*, *FNDC5*, *CXCL1*, *SPARC*, *IL6*, and *IL15*; and downregulation of *FGF21*. This data may represent that the overexpression of almost all of these genes can be contributed to their main roles as anti-cancer effects based on previously reported studies.

Our results revealed that the most notable significant alterations for them at each stage were as follow: D2 (*IL10* and *CXCL1*), D7 (*FGF21* and *DCN*), D30 (*FGF21*, *DCN*, *MSTN*, and *FNDC5*), D46 (*FGF21*, *DCN*, *MSTN*, and *FNDC5*) (Figure 35). *SPARC* and *DCN* showed high significant overexpression, in order, from D7 upward. Moreover, IL-10 revealed overexpression only on D2. Furthermore, *IL15* and *MSTN* showed overexpression on D30 and D46 while *BDNF* overexpression was observed on D2 and D30. Moreover, *FNDC5* showed overexpression on D2, D30, and D46. Also, *IL6* and *CXCL1* showed overexpression on D46 and their lower expression on D2. *FGF21* was the only gene in this category that showed significant downregulation at all stages compared to D0 (Figure 35).



Figure 35. Altered expression of myokines across muscle differentiation. The comparison between muscle differentiation stages and undifferentiated cells (D0) is based on CPM and their significant FDR value for two upper bar plots and based on logFC and FDR for lower bar plot. *: p-value less than 0.05; **: p-value less than 0.01; ***: p-value less than 0.001.

3.5 Involvement of breast cancer genes in muscle differentiation

Using GDCRNATool, we extracted various genes with altered expression affecting OS patients from TCGA breast cancer (Table S25) and then we used these genes to investigate across different stages of muscle differentiation. Our results revealed that a number of these genes are involved in muscle differentiation (Figure S10-12 and Table S26). Some of these genes had altered expression at all stages compared to D0 as follow (Figure S10 and Table S26). The genes with lower expression were *MAL2*, *GRHL2*, *CD24*, *FAM155B*, *EPB41L4B*, *TPD52*, *ESRP1*, *SUSD3*, *KLK10*, *DNMT3B*, *KIF17*, *STXBP1*, *SYN2*, *KRTCAP3*, *CLDN7*, and *XG*. However, the genes with higher expression included *DIRAS3*, *LIMCH1*, *WLS*, *FREMI1*, *C2orf88*, *AFAP1L2*, *RASSF9*, *LEF1*, *PDLIM4*, *ACSL5*, *CTHRC1*, *PPP1R14C*, *COL3A1*, and *PCDH17* (Figure S10 and Table S26). Among TCGA BRCA genes with altered expression at all muscle differentiation stages compared to D0, *HPGD*, *NR3C2*, *ALDH1A1*, *SORCS1*, and *FER1L4* revealed lower expression at first stages (D2 and 7) but higher expression at late stages (D30 and 40). Moreover, only *EMP1*, *BOC*, and *CRYAB* showed downregulation on D2 but upregulation at other stages compared to D0 (Figure S10 and Table S26). However, some genes showed expression changes at specific stage or stages but not at all stages given in Figures S11 and 12.

Among the genes with significant OS from TCGA breast cancer, we found several uncharacterized and novels ones with higher expression (*SLC35A2*, *DONSON*, *BRI3BP*, *NDUFAF6*, *C2CD4D*, and *SEZ6L2*) (Table 6) and lower expression (*N4BP2L1*, *BOC*, *RIC3*, *XG*, *BEGAIN*, and *PLEKHA4*) in TCGA-BRCA (Table 7). However, here, we mainly focus on uncharacterized and novels ones with possible tumor suppressor role.

Table 6. Uncharacterized genes affecting OS with possible new proto-oncogene roles in breast cancer.

Symbol	HR	lower95	upper95	pValue	logFC	PValue	FDR
SLC35A2	1,63	1,186	2,25	0,0025	1,28	6,69E-90	2,40E-88
NDUFAF6	1,6	1,16	2,19	0,0047	1,04	1,08E-31	6,30E-31
LINC01094	1,51	1,1	2,09	0,01	1,05	4,95e-26	2,29e-25
AC055854.1	0,66	0,48	0,91	0,013	2,096	1,8e-09	3,77e-09
AC124798.1	0,67	0,49	0,92	0,013	1,11	2,28e-15	6,49e-15
AC098934.1	1,50	1,09	2,06	0,014	2,08	1,85e-30	1,027e-29
LINC01977	1,43	1,04	1,97	0,026	1,7	6,45e-18	2,06e-17
CYP4F23P	0,70	0,51	0,97	0,031	1,58	3,03e-08	5,86e-08
SEZ6L2	1,39	1,008	1,91	0,044	1,129	1,81E-08	3,54E-08
DONSON	1,39	1,007	1,91	0,046	1,21	2,24E-39	1,71E-38
BRI3BP	1,38	1,002	1,90	0,047	1,021	8,77E-38	6,32E-37
C2CD4D	0,73	0,53	0,99	0,048	1,186	8,88E-21	3,25E-20

Table 7. Down regulation of TCGA BRCA genes affecting OS functioning as possible new tumor suppressors in breast cancer.

Symbol	HR	lower95	upper95	pValue	logFC	PValue	FDR
RIC3	0,73	0,53	1,00	0,049	-3,022	3,14E-64	5,66E-63
BEGAIN	0,69	0,50	0,95	0,023	-1,406	8,85E-28	4,39E-27
BOC	0,72	0,52	0,99	0,042	-2,221	1,82E-81	5,45E-80
XG	1,39	1,006	1,91	0,044	-1,374	4,33E-28	2,17E-27
N4BP2L1	0,73	0,53	0,99	0,047	-1,46	7,72E-91	2,84E-89
PLEKHA4	0,61	0,44	0,84	0,0028	-1,2	1,40E-27	6,87E-27
LINC00987	0,69	0,50	0,96	0,024	-1,88	3,57e-77	9,62e-76
AC109587.1	0,71	0,51	0,98	0,033	-1,16	8,66e-73	2,038e-71
AC105942.1	0,50	0,50	0,96	0,029	-1,08	3,76e-29	1,97e-28

Using Enrichr, we found that the uncharacterized novel tumor suppressor genes from TCGA-BRCA had directly or indirectly roles in myogenesis based on Enrichr GO:BP (Figure 36 and Table S27). This myogenesis involvement is mainly for *BOC* gene (positive regulation of myoblast differentiation (GO:0045663), positive regulation of muscle cell differentiation (GO:0051149), regulation of myoblast differentiation (GO:0045661), regulation of muscle cell differentiation (GO:0051147)) and *PLEKHA4* involved in regulation of Wnt signaling pathway which is critical for induction of mesoderm lineage (positive regulation of Wnt signaling pathway, planar cell polarity pathway (GO:2000096)), positive regulation of non-canonical Wnt signaling pathway (GO:2000052), positive regulation of canonical Wnt signaling pathway (GO:0090263) (Figure 36 and Table S27). Moreover, based on Enrichr, *BEGAIN* gene is involved in muscle differentiation since it has function in regulation of postsynaptic neurotransmitter receptor activity (GO:0098962 and GO:0099601) and modulation of chemical synaptic transmission (GO:0050804) (Figure 36 and Table S27), with important role in neuromuscular junctions (NMJs), a cellular synapse between a motor neuron and a skeletal muscle fiber [65]. Looking at our muscle differentiation data, it also showed their direct involvements in muscle differentiation (*PLEKHA4* and *BEGAIN* revealed significant upregulation only on D30 and D46 but *BOC* showed downregulation on D2 while upregulation on D7, 30, and 46) (Figure 37A and B). Furthermore, based on Enrichr, *XG* seems to be involved in muscle differentiation since it is involved in homotypic cell-cell adhesion (GO:0034109) (Figure 36 and Table S27), which is important during mesoderm lineage and other stages of muscle differentiation [66]. In our muscle differentiation, it also showed significant downregulation at all stages (D2, 7, 30, 46) compared to D0 (Figure 37A and B).

RIC3, based on Enrichr, has key role in regulation of protein localization to cell surface and synaptic transmission, cholinergic (positive regulation of protein localization to cell surface (GO:2000010), synaptic transmission, cholinergic (GO:0007271), protein localization to cell surface (GO:0034394), regulation of protein localization to cell surface (GO:2000008), and positive regulation of cellular protein localization (GO:1903829)) (Figure 36 and Table S27). These processes are important in

muscle lineage and looking at our muscle differentiation stages, *RIC3* had significant downregulation only at late differentiated stages (D30 and D46) compared to stemness stage (Figure 37A and B). Regarding *N4BP2L1*, while based on Enrichr it did not show involvement in any biological processes (Figure 36 and Table S27), it is reported to be involved in Inflammatory Leiomyosarcoma (<https://www.genecards.org>) which is originated from muscle tissues and our muscle differentiation revealed its significant overexpression at late stages (D30 and D46) (Figure 37A and B). Therefore, for the first time, we proposed its role in muscle differentiation and also breast cancer. Altogether, we found an interesting overexpression of these genes, mainly three genes with direct involvement in muscle differentiation (*PLEKHA4*, *N4BPL2L*, and *BEGAIN* revealed upregulation only at late differentiated stages, D30 and D46), and *BOC* with upregulation on D7, 30, and 46 but down regulation in D2. In this category, the main downregulated genes across the differentiation stages were *XG* (at all stages: D2, 7, 30, and 46) and *RIC3* (on D30 and D46) (Figure 37A and B). Regarding uncharacterized and possible novel proto-oncogenes from TCGA BRCA, main significant downregulation, as a characteristic of proto-oncogenes in differentiated cells, was for *SEZ6L2* on D2, D7, and D46; and *BRI3BP* on D30 (Figure 37C).

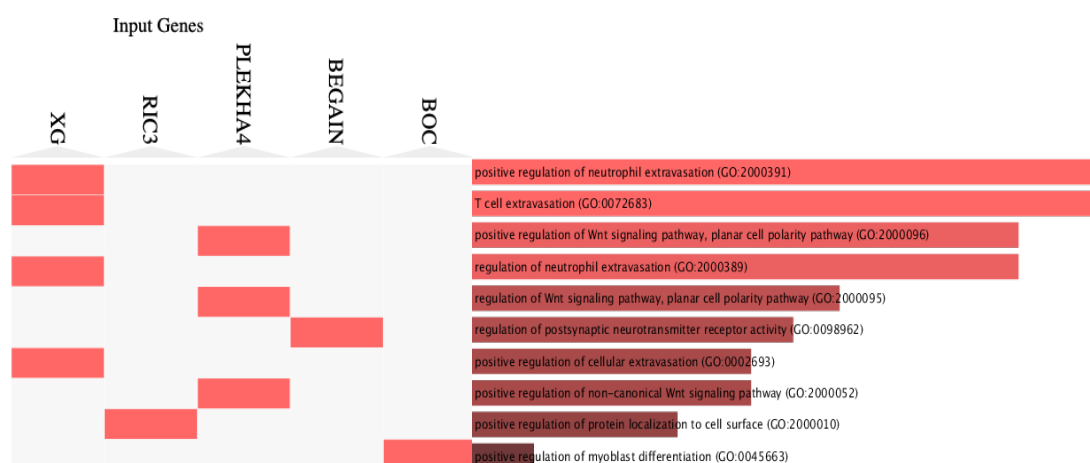
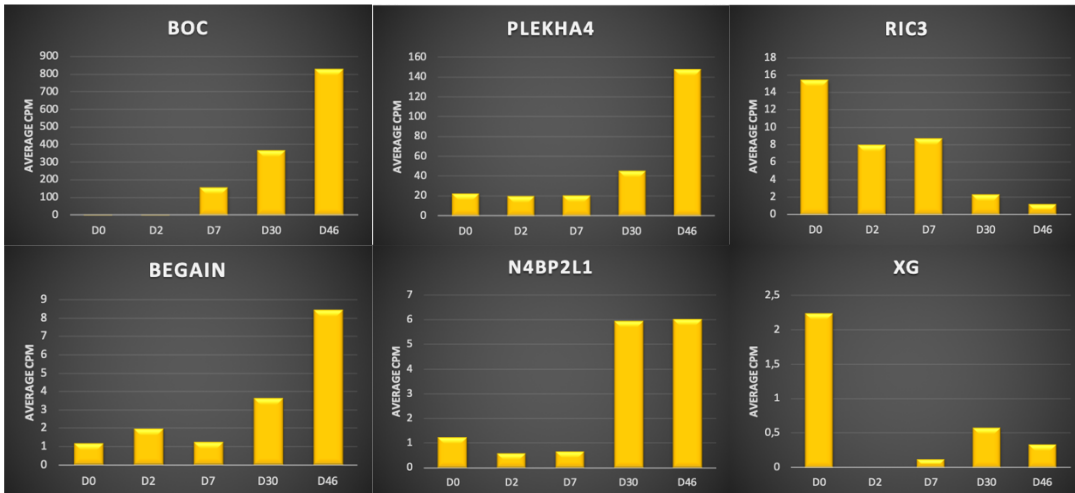
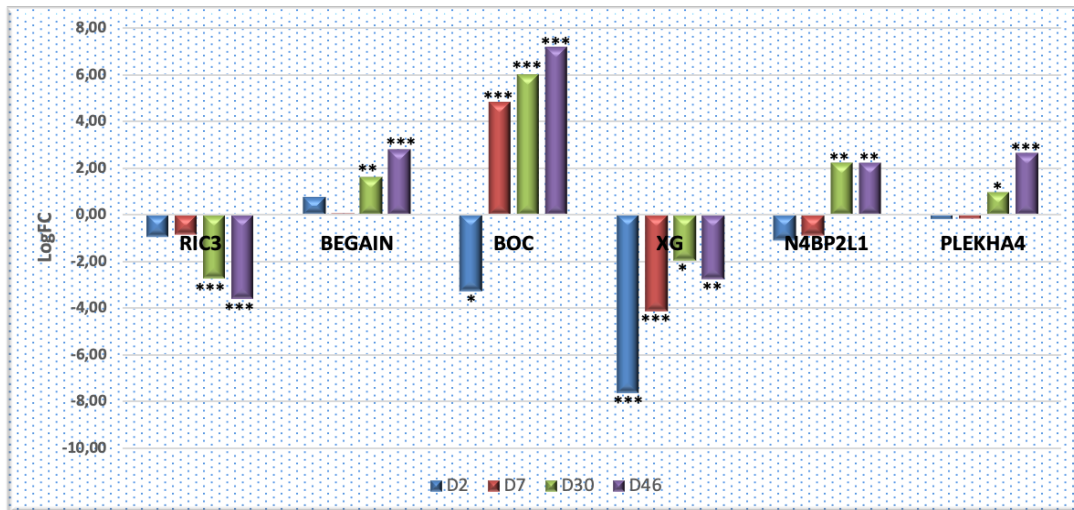


Figure 36. Enrichr GO:BP terms for the novel tumor suppressor genes from TCGA-BRCA. GO:BP terms are listed based on adjusted p-value (the value less than 0.05 were considered statistically significant.).

A



B



C

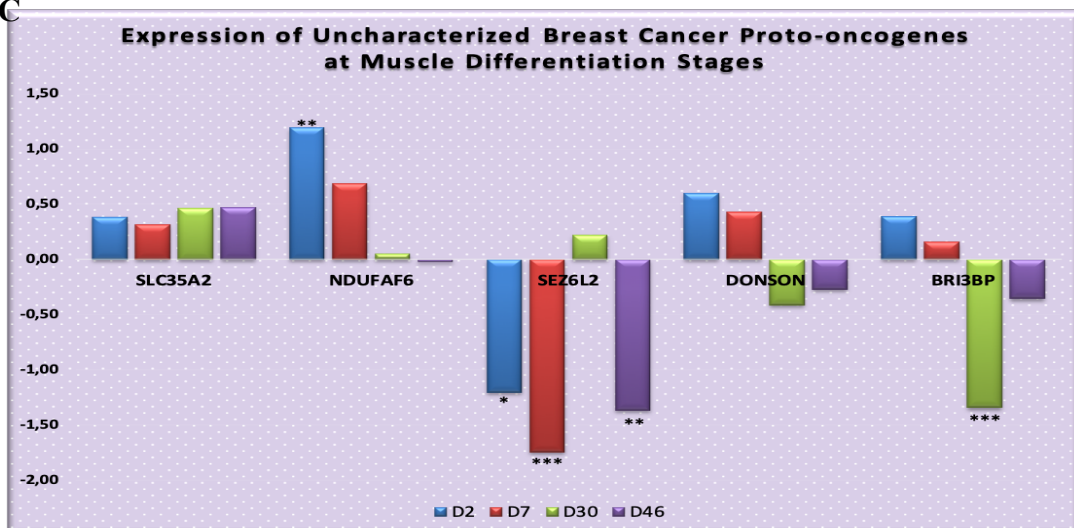


Figure 37. (A) Altered expression of novel tumor suppressors based on CPM and their significant FDR compared to D0. **(B).** Significant expression of the novel tumor suppressors is shown separately based on logFC compared to D0. **(C)** Significant expression of Proto-oncogenes from TCGA-BRCA across muscle differentiation stages compared to D0. *: p-value less than 0.05; **: p-value less than 0.01; ***: p-value less than 0.001.

3.6 Involvement of ubiquitin proteins in muscle differentiation and cancer

As mentioned previously in different sections, pectoralis muscle from breast cancer patients had an elevated expression of genes involved in ubiquitin-mediated proteolysis [35, 57]. Therefore, we investigated the expression of ubiquitin proteins and their associated proteins across muscle differentiation stages.

Regarding E2-conjugating enzymes and some of their associated proteins, we found the significant altered expression of a number of genes. Some of these genes had higher expression in the stemness condition compared to muscle cell progenitor stage (D30) and/or terminal stage (D46) such as *UBALDI* (lower expression at all stages), *UBE2O*, *UBE2S*, *UBA2*, *UBE2C*, *UBE2T* (lower expression on D30 and D40), *UBE2N* and *UBE2V2* (significant lower expression only on D46). *UBE2H* and *UBE2L6* showed only lower expression on D2 (Figure 38A and B). However, some genes had higher expression across this muscle process; For example, significant higher expression of *UBA7*, *UBASH3B*, *UBAP1L*, and *UBE2D4* at late stages (D30 and D46); *UBE2QL1* on D30; *UBE2E3* and *UBE2J1* on D7, 30, and D46; and *UBTD2* on D7 and D46 (Figure 38A and B).

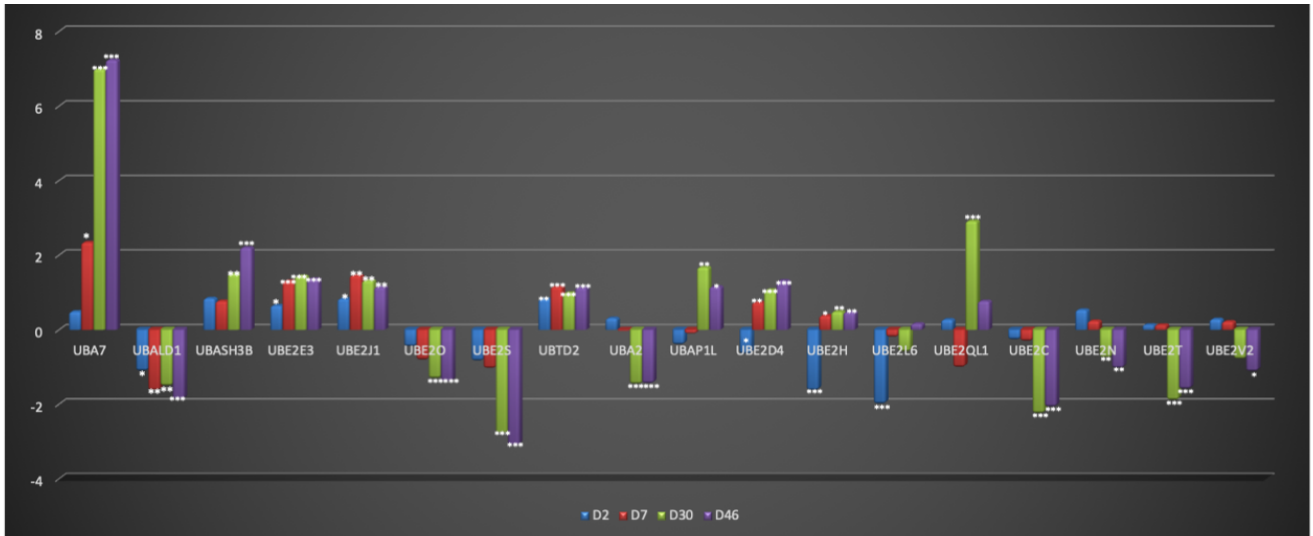
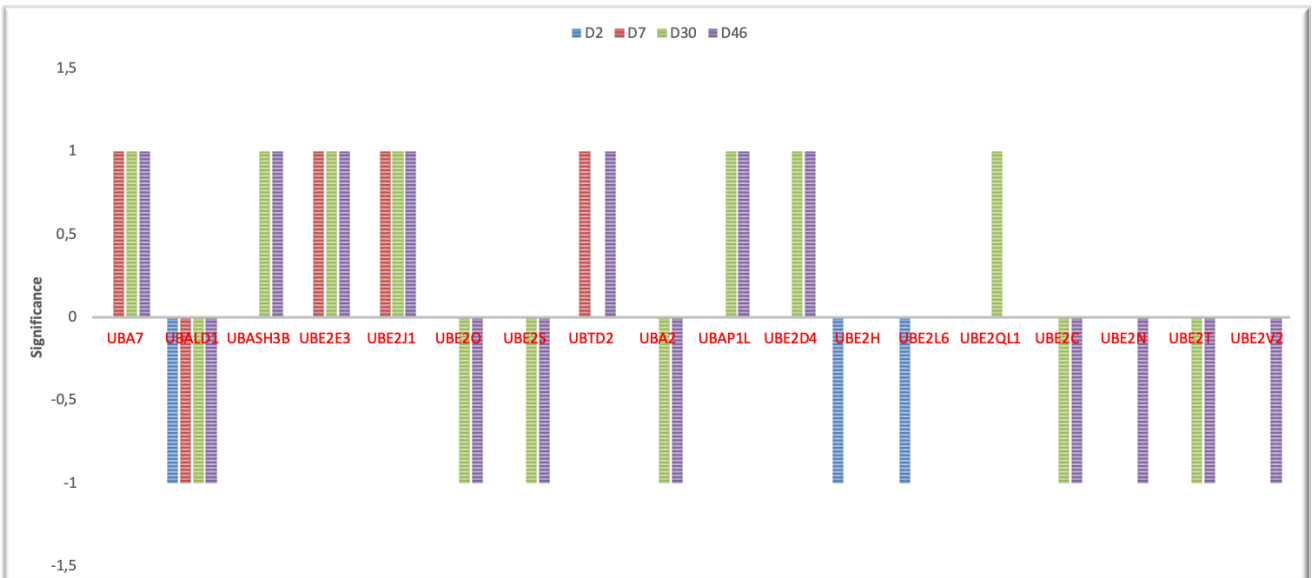
A**B**

Figure 38. Altered expression of some important E2-conjugating enzymes and their associated proteins across muscle differentiation stages compared to D0. **A.** The expression pattern based on count per million (CPM) compared to D0. **B.** The expression pattern is based on significance of expression alteration (+1 for over expression and -1 for lower expression) compared to D0. *: p-value less than 0.05; **: p-value less than 0.01; ***: p-value less than 0.001.

In relation to the ubiquitin ligases, some genes revealed significant expression changes at all muscle differentiation stages with lower expression for *CDH1*, *HECW1*, *HERC5*, and *RNF125* and overexpression for *HECTD2* and *HECW2* (Figure 39A and B). Some genes only showed significant changes at both late stages (D30 and D46) with lower expression for *CDC20*, *DTL*, *RNF138*, *RNF168*, *RNF5*, and *UBE3D*; and higher expression of *PJA2* and *TRIM63* (Figure 39A and B). However, some genes showed lower or higher expression specifically on both D2 and D7 with lower expression of *LNXI*, *MARCH2*, and *RNF128*; and higher expression of *MAP3K1* and *NEDD4* (Figure 39A and B). In addition to these changes, some of these genes showed significant altered expression at specific stages [*NEDD4L* (↓) and *WWP1* (↑) on D2; *RFFL* (↓) on D7; *MARCHE1* (↑), *MARCHE6* (↑), and *TOPRS* (↓) on D30; and *CBL* (↓) and *PEL1I*(↓) on D46; *MARCH4* (↓) on D2 and D46]. Moreover, other combinations can be observed (*MARCH3* (↓) on D2 and (↑) on D30). Furthermore, *PRKN* revealed overexpression on D7, 30 and 46 (Figure 39A and B).

However, several ubiquitin enzymes and their associated proteins did not show significant altered expression during muscle differentiation, mainly interestingly *UBAPI* which showed a constant expression at all stages (Figure 40, average CPM as follow: D0: 54.14; D2: 52.4; D7: 55.49; D30: 56.45; D46: 52.06). For the first time in collaboration with Persian BayanGene Research and Training Center in Iran and also University of Miami, we found that truncating mutations in *UBAPI* resulted in the hereditary spastic paraplegia (the Spastic paraplegia 80, autosomal dominant), a rare inherited disorder of weakness and stiffness in the leg muscles [67]. The *UBAPI* gene has instruction for producing a component of the endosomal sorting complex required for transport-1 (ESCRT-I) and has a key role in proteasomal degradation of ubiquitinated cell-surface proteins. Therefore, its constant expression may reveal its important role during all differentiation stages.

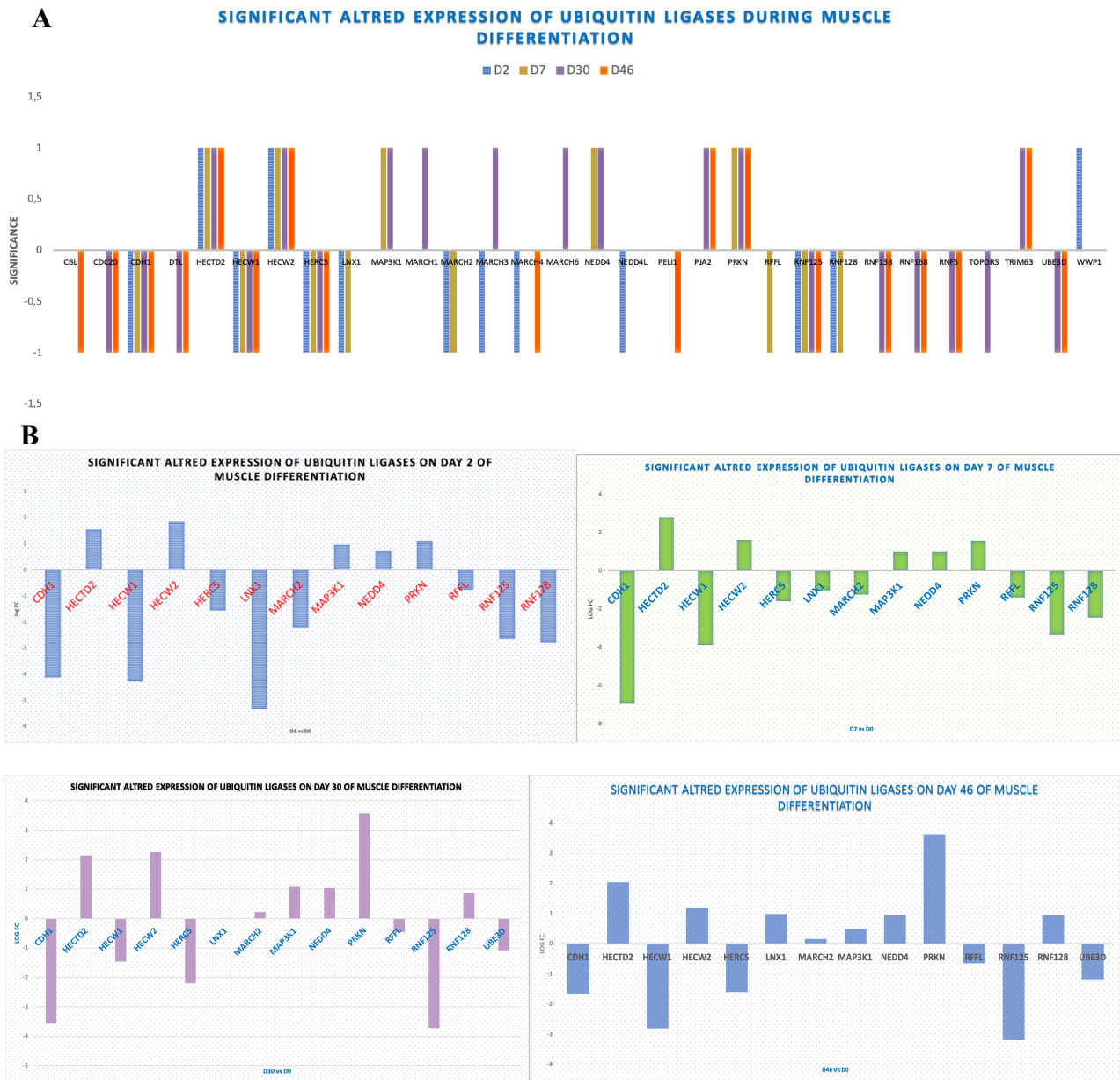


Figure 39. Altered expression of ubiquitin ligases across muscle differentiation stages. **A.** The comparison is based on significance of expression (+1 for overexpression and -1 for lower expression) compared to D0. **B.** The comparison is based on logFC and their significant FDR value compared to D0.

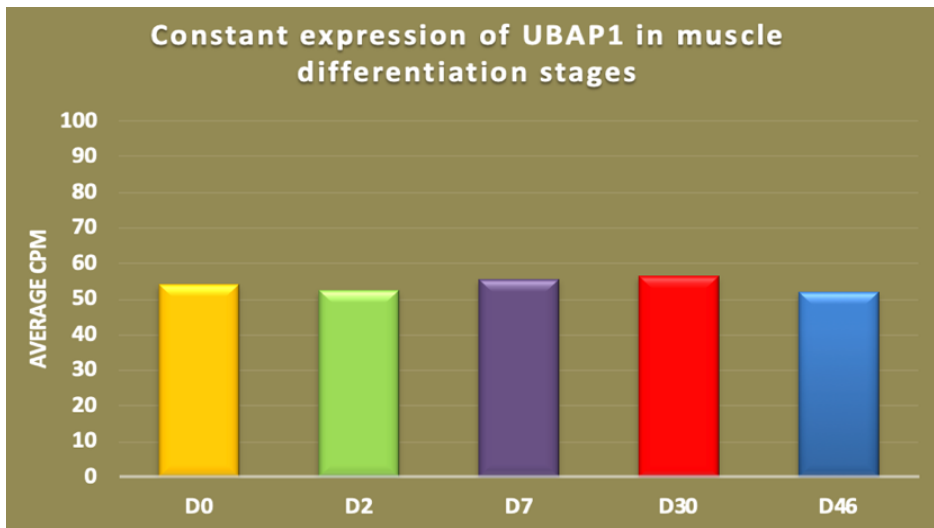


Figure 40. *UBAP1* expression across muscle differentiation stages. The comparison is based on average CPM.

3.7 *UBE2C* involvement in tumors and muscle differentiation

As above-mentioned sections, ubiquitin proteins are mainly involved in muscle differentiation stages. We showed that *UBE2C* had significant lower expression at late stages compared to stemness stage, indicating its function as proto-oncogene.

Our previously reported study showed that *UBE2C* expression was notably higher in all 27 TCGA tumor types compared to their matched TCGA normal tissues and GTEx data [36]. Moreover, our study revealed that the expression of *UBE2C* had a moderate to very strong positive correlation with other genes in 27 TCGA cancers [36].

The overexpression of *UBE2C* in cancers, including breast tumor, can confirm its key involvement in the stemness status of the cancer since *UBE2C* has higher expression at stemness stage of muscle differentiation compared to terminal differentiated stages. In this part of our study, we analyze the expression of *UBE2C* and its significant positively co-expressed genes from TCGA cancers, mentioned in method section [36], during muscle differentiation stages.

It is worth to note that our results found the parallel lower expression of all these genes, except *CDK1* and *BRCA1* genes, along with *UBE2C* at two late stages (D30 and D46) of muscle differentiation

stages, representing their possible role as proto-oncogene (Figure 41 and Table S28). These data may strengthen their involvement in the same networks in cancers and stemness of cancer since in muscle differentiation, they showed lower expression at late differentiated stages compared to stemness stages. Enrichr GO:BP terms revealed some important expected processes for these genes (Figure S13 and Table S29) such as regulation of mitotic cell cycle phase transition (GO:1901990), anaphase-promoting complex-dependent catabolic process (GO:0031145), regulation of cell cycle phase transition (GO:1901987), regulation of exit from mitosis (GO:0007096), positive regulation of ubiquitin protein ligase activity (GO:1904668), mitotic spindle organization (GO:0007052), regulation of mitotic cell cycle (GO:0007346), mitotic cell cycle phase transition (GO:0044772), regulation of ubiquitin protein ligase activity (GO:1904666), and positive regulation of ubiquitin-protein transferase activity (GO:0051443).

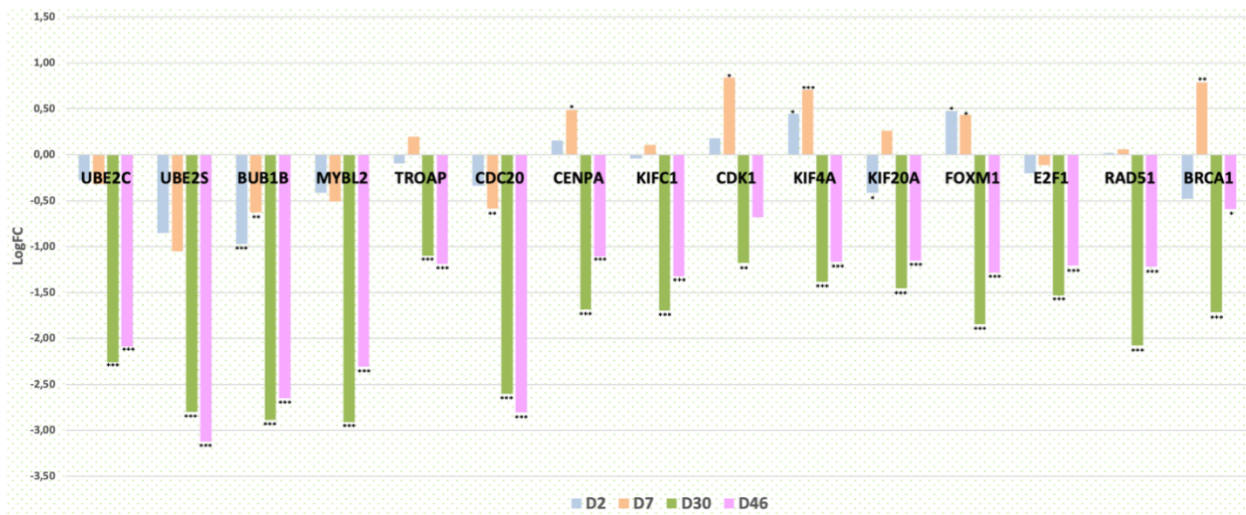


Figure 41. *UBE2C* positively co-expressed genes and from TCGA BRCA show altered gene expression across muscle differentiation. The comparison for part A is based on logFC and their significant FDR value compared to D0. *: p-value less than 0.05; **: p-value less than 0.01; ***: p-value less than 0.001.

Chapter 4

Discussion

hESCs as a hPSCs type have remarkable capability to indefinitely self-renew in the undifferentiated state and the capacity to differentiate into all three embryonic germ layers (Figure 1)[1, 2]. These cells have an enormous potential as a source for cell-based therapies and disease model systems [1, 4]. Stem cell biology advances have been increasing in the last decade and blossomed into a very promising and advanced scientific research field; for instance, their use in disease treatment and generation of neuromuscular disease models, and identification of biological pathways across various cell differentiation processes [12]. Recently, there has been a great interest to use myogenic progenitors differentiated from ES cells for regenerative medicine [8, 13]. In addition, since recent studies have provided the involvement of skeletal muscle cells in cancer [14-17] [34], interests for the identification of the pathways related to differentiation of hESCs to muscle cells have been interestingly increased.

Therefore, in our study we provided a well described protocol developed by Shelton, M et al. [4] (Figure 2) to differentiate H9 hESCs into muscle cell and its progenitors (PAX7+ expressing SMPs). The presence of these SMPs along with myocytes can make its therapeutic advantages since the dividing SMPs help differentiate them toward new myocytes and maintain the PAX7+ expressing SMPs pool [4]. Using this protocol, we could successfully differentiate H9 hESCs to muscle cell and its SMPs for 46 days and helped uncover pathways and genes with/out proto-oncogenic, tumor suppressive, TF, and Epifactors roles during different muscle differentiation stages. As expected, we confirmed the process with confirmation of high expression of stage markers of Brachyury (*T*) on D2 (mesoderm stage), *PAX3* on D7 (as somite marker), *PAX7* on D20 (preferential expansion of SMPs cells) and also the expression of *MYOG* on D30 (the early dermomyotome-like stage), and finally the *MYH3* (*MF20*, myosin heavy chain, as terminal muscle differentiated marker) on D46 and also presence of PAX7+expressing SMPs.

Our RNA sequencing data revealed high quality of samples in terms of read counts, distribution and presence of stage markers. Our RNA sequencing results validated the expected presence and high expression of specific markers for each stage (Figure 13). The H9 hESCs presented optimal expression of embryonic stem cell markers *POU5F1*, *SOX2*, and *NANOG* (D0). Then, at paraxial mesoderm stage (D2), the cells had the high expression of *TBXT*, *MSGN1*, and *TBX6*. Higher expression of *PAX3*, *MEOX1*, *MEOX2*, and also *PAX7* on D7 confirmed the efficient somite stage. Furthermore, SMPs markers (*MYOG*, *MYOD1*, and *PAX7*) revealed elevated expression on D30 to confirm the presence of myogenic progenitors and *PAX7*⁺ expressing SMPs. At final stage of the differentiation, the higher expression of genes with myosin heavy chain such as *MYH7* and also *MYOG* were confirmed. Moreover, at this terminal stage, the expression of *PAX7* was observed. It is worth to note that our study is among the few studies which confirms the presence of *PAX7* expression along with *PAX3* at somite stage [6, 62, 63] and also the expression of *SOX2* as expected for somite stage [64].

Overall, our differentiation process revealed a number of differential expressed genes which grouped into nine clusters of gene expression with the clearly distinct patterns of regulation over the muscle differentiation stages. The gene clusters were grouped into stem stage, early stage, middle stage, late stage and late & stem stage based on the gene expression in clusters (Figure 14).

Our cluster profiling for GO: BP identified different terms for each muscle differentiation stage (Figure 15) which were in concordance with expected behaviors of each stage. For instance, GO: BP with muscle cell development was specific for the late stage (Figure 15), among others, explained in results section. Furthermore, our KEGG pathways profiling revealed essential pathways for each stage/s (Figure 16). For example, activation of Wnt signaling pathway is the main one for differentiation of stem cells toward muscle lineage which is achieved using GSK3 $\alpha\beta$ inhibition. Several of these pathways are also involved in cancers such as PI3K-Akt signaling pathway, Rap1 signaling pathways, Wnt signaling pathway, among others.

In our transcriptome profiling, we then investigated the expression of genes with different functions such as TF, Epifactor, proto-oncogene, and tumor suppressor; and TCGA-BRCA genes, genes involved in ubiquitin system, myokines, among others.

Our differentially expressed gene analysis across the muscle stages revealed a notable altered expression of a various number of genes at all stages or at a specific stage/s. It should be noted that our data analysis revealed altered expression of a number of transcription factors, epigenetic factors, and lncRNAs (Figure 17). Moreover, our data showed involvement of numerous tumor suppressors and canonical cancer drivers found in specific clusters (Figure 18A).

In relation to the cancer drivers as proto-oncogenes and tumor suppressors, our data revealed the significant downregulation of several genes at all stages or a specific stage/s (Figure 18B) compared to stem stage (D0). Among this category, the most significant downregulated gene was *POU5F1* (a TF functioning as proto-oncogene), *USP44* (an Epifactor functioning as proto-oncogene), *LCK* (a gene without TF and Epifactor roles functioning as proto-oncogene) among others, indicating their roles in the stemness status of hESCs and cancer cells. In this category, we also found one of the main proto-oncogenes involved in sarcoma, as a muscle-related cancer, (*BCOR*). This gene shows overexpression in sarcoma affected by BCOR-ITD [68] and dysregulated in fusions with other genes [69]. In our muscle differentiation stage, significant *BCOR* downregulation was observed at stages of D7, 30, and 46 compared to D0, confirming its function as proto-oncogene and its role in stemness of hESCs.

In our study, we also found some cancer drivers with only TF role functioning as proto-oncogene, with more significant altered expressions at all stages for lower expression of *POU5F1* and *MYC*; and overexpression of *LEF1* and *PAX3* compared to D0. However, other important ones were *MYCN* with lower expression only on D2, D30, and D46; and *MYB* with lower expression on D30 and 46 (Figure 20), indicating their role in stemness status of hESCs. These genes are involved in several cancers [70-74].

Among genes with only Epifactor role with proto-oncogenic function, almost all of the genes in this category were downregulated across almost all stages (Figure 21), obviously representing their involvement in stemness status of stem cells and cancers.

In relation to the genes with both TF and Epifactor roles and functioning as both proto-oncogenes and tumor suppressors, the most significant alteration was for *ZBTB16* with overexpression from D7 toward late stage. *ZBTB16* is reported to be involved in Leukemia, acute promyelocytic, PL2F/RARA type (<https://www.omim.org/>) [75]. Other genes were *EZH2* (significant higher expression at mesoderm and somite stages), *PBRM1* (overexpression at somite stage), and *TP53* (significant lower expression at late stages of differentiated) (Figure 22). Moreover, among genes with both proto-oncogenic and tumor suppressive functions but only as TF, our data showed a significant lower expression of *FOXL2* from D7 toward late stage compared to D0 (Figure 23). In relation to the cancer drivers with only TF role which act as tumor suppressors, we identified significant overexpression of *GATA4* and lower expression of *SALL2* at all stages (Figure S1). Moreover, our data revealed altered expression of genes defined as Epifactors with tumor suppressive function, in which almost all of them had lower expression across muscle differentiation (Figure 24). The important gene in this category was *DNMT3B* with very significant lower expression at all stages compared to stemness stage. Furthermore, two cancer drivers which are considered as both TF and Epifactor with tumor suppressive function, were *CTCF* with lower expression at all stages and *SMARCC1* with lower expression on D30 and 46 (Figure S2).

Regarding tumor suppressor genes with TF role, interestingly almost all of them revealed higher expression at almost all muscle differentiation stages compared to D0. The most significant values were for *IFI16* and *FOXCI* across all stages (Figure S3), indicating their important role in differentiated cells as expected for tumor suppressor. In relation to the tumor suppressor genes with only Epifactor role, most of them were upregulated. The main significant altered genes with higher expression at all stages compared to stemness stage (D0) were *SMARCA2*, and *EYA4* and also *SIRT2* with higher expression from D7 to D46 (Figure 25). This represents their key roles in differentiated

cells compared to stemness status. Moreover, among genes with role of both TF and Epifactors with tumor suppressive function, our data revealed altered expressed of some genes. The main significant one was *SP100* overexpression from D7 toward terminal stage compared to D0 as expected for tumor suppressors (Figure 26).

In addition to the above-mentioned data, looking at whole data, our results showed the altered expression of some genes with both TF and Epifactor roles functioning as both proto-oncogenes and tumor suppressors, with more significant alterations for *FOXO1* downregulation at all stages and *ZBTB16*, *FOXP1* and *PRDMI* upregulation from D7 toward late stage compared to D0 (Figure 27).

Regarding proto-oncogenes without TF and Epifactor roles, our data found major significant alterations for downregulation of *LCK*, *BUB1B*, *RMI2*, and *KDR* at all stages and *RHOH* from D7 toward late stage as expected for proto-oncogenes; and overexpression of *COL2A1* at all stages (Figure 28). However, in relation to the tumor suppressor genes without TF and Epifactor role, interestingly, we identified a number of overexpressed genes across muscle differentiation as expected for these genes (Figure 29). The most significant changes were for *EMPI* with upregulation from D7 toward the late stage. Other main significant altered genes were *PCDH9*, *VIM*, *WNT5A*, and *AHNAK* with overexpression at all stages; and *GAS1*, *DMD*, and *IGFBPS* with overexpression from D7 toward the late stage.

In another category with tumor suppressors and proto-oncogenes without TF and Epifactor roles among all clusters, our results revealed several main genes, in which most of them had upregulation during muscle differentiation stages (Figure S4). The most significant alteration was for *EPHA3* with upregulation from D7 toward the late stage compared to D0. Among these genes in this category, some important known genes are involved in cancers such as *CDH11*, *CDKN2C*, *DCC*, *LRIG3*, *DDR2*, *LRP1B*, *CDK6*, and *CDKN2A* with overexpression; and *CDH1* and *WNK2* with lower expression at all stages.

Regarding the involvement of genes with role as both TF and Epifactor extracted from all clusters of muscle differentiation stages, we identified several genes with altered expression at all stages or

specific stage/s. The major significant alterations were for downregulation of *CTCF*, *FOXO1*, *PRDM14*; and upregulation of *DPF3*, *PRDM8*, *SCML4*, *SNAI2*, *ZHX1*, and *ZNF516* at all stages compared to D0 (Figure S5 and 30). Their main GO:BP terms were regulation of transcription (Figure 31A).

We also found overexpression of some *FOX* genes from D7 toward terminal differentiation, representing their important role in muscle differentiation with their function in the regulation of transcription with involvement of all *FOXA1*, *FOXP1*, *FOXP2*, and *FOXP4* genes (Figure S6B and Table S15).

Looking for the involvement of DNMTs with Epifactor role across muscle differentiation stages, we found significant *DNMT3B* lower expression at all stages compared stemness stage. However, *DNMT3A* lower expression was observed only at mesoderm stage and *DNMT1* down regulation only at late stages of D30 and 46 (Figure 32). This finding may represent the key role of *DNMT3B* in stemness status compared to differentiated cells and also confirm its role in stemness of cancers since in several TCGA cancer *DNMT3B* overexpression is found. Moreover, the data may indicate that *DNMT3A* expression is important in differentiated cells. In category for the genes with only Epifactor role, we identified a number of genes with altered expression across muscle differentiation stages (Figure S7 and S8) and their main Enrichr GO:BP term was regulation of transcription and the major enriched genes were *SMARCA2*, *RBI*, and *CHD3*, in order (Figure 33A).

In one part of our investigations, we found the involvement of main myokines in muscle differentiation, with their higher expression across this process. Our data revealed that the main significant alterations were for overexpression of *DCN*, *MSTN*, *FND5*, *CXCL1*, *SPARC*, *IL6*, and *IL15*, and downregulation of *FGF21* (Figure 35). Therefore, their higher expression at differentiated stages and also their release after physical activities with anti-cancer effects may confirm the fact that they have role in reduction of stemness status of cancer and stem cells.

In another part of our study, based on the fact that pectoralis muscle mass enhancement has been found in samples with breast tumors [34] and also these tissues in patients with breast tumors

represented altered expression of genes involved in muscle cell homeostasis [35, 57], we investigated and observed the involvement of TCGA BRCA genes affecting OS across muscle differentiation. Our results showed involvement of a number of these genes in muscle differentiation (Figure S10-S12 and Table S26). Some of these genes had altered expression at all stages compared to D0 with possible functions as proto-oncogene and/or tumor suppressor (Figure S10 and Table S26). Moreover, we showed involvement of some uncharacterized and novel TCGA BRCA tumor suppressors (*N4BP2L1*, *BOC*, *RIC3*, *XG*, *BEGAIN*, *PLEKHA4*) at muscle differentiation stages. This involvement is mainly considered for *N4BP2L1*, *BOC*, *N4BP2L1*, and *BEGAIN* based on their significant overexpression at late muscle stages and also their involvement in muscle related processes extracted from Enrichr GO:BP (Figure 36 and Table S27) [65] [66]. Briefly, *BOC*, *PLEKHA4*, *BEGAIN*, and *N4BP2L1* with higher significant expression at late muscle stage may strengthen their key function as tumor suppressor. Overall, we could confirm the role of these breast cancer genes in muscle differentiation and for the first time, we presented the role of the *PLEKHA4*, *BEGAIN*, *XG*, *RIC3* and *N4BP2L1* in muscle differentiation.

Regarding the involvement of ubiquitin proteins in muscle differentiation and cancers, we found a significant altered expression of a number of genes with higher expression in stemness condition compared to muscle cell progenitor stage (D30) and/or terminal stage (D46). For example, significant lower expression of *UBE2S* and *UBE2C* on both D30 and D40 compared to stemness stage D0 (Figure 38A and B; and Figure 39). This may indicate their main roles in the maintenance of hESCs stemness. In addition to this finding, various ubiquitin proteins did not have significant changes in their expression at muscle differentiation stages compared to D0. One of the main ubiquitin proteins with constant expression at all muscle stages was *UBAPI* (Figure 40). For the first time in collaboration with other research groups, we showed that the impaired *UBAPI* could cause the hereditary spastic paraplegia (the Spastic paraplegia 80, autosomal dominant) which is a rare inherited disorder with weakness and stiffness in the leg muscles [67]. Ubap1 protein has a vital role in proteasomal

degradation of ubiquitinated cell-surface proteins and its constant expression may reveal its important role during all differentiation stages.

In the last part of our study, we could show the involvement of *UBE2C* and its co-expressed genes from TCGA cancers at muscle differentiation stages. In our previous published study, we identified *UBE2C* overexpression in all 27 TCGA cancers and its expression had a moderate to very strong positive correlation with various genes in the twenty-seven cancers. In addition to *UBE2C* involvement in cancer, as mentioned above, we identified *UBE2C* significant lower expression at late stages of muscle differentiation compared to stemness stage, indicating its function as proto-oncogene. The overexpression of *UBE2C* in cancers [36], mainly breast cancer, and its higher expression at stemness stage of muscle differentiation compared to terminal differentiated stages can confirm its key involvement in the stemness status of the cancer and hESCs. Furthermore, when we analyzed the expression of *UBE2C* and its positively significant co-expressed genes from TCGA cancers during muscle differentiation stages, our result revealed the parallel lower expression of almost all these genes along with *UBE2C* at two late muscle stages (D30 and D46), representing their possible role as proto-oncogenes (Figure 41). These data may confirm their involvement in the same *UBE2C* networks in cancers and its stemness.

In summary, our hESCs differentiation toward muscle cells could help uncover a huge number of genes involved in this cell process contributed to the most important human body tissue. Most of them can be considered for our future functional studies to identify their exact roles in this process. Several genes showed roles as TF and/or Epifactors which also acted as proto-oncogene and/or tumor-suppressor. We could propose several genes as new proto-oncogenes and tumor suppressors. Moreover, we could confirm the direct and indirect involvement of several genes with few reports for their involvement in different stages of myogenesis such as somite stage. We were also able to report lots of genes involved in cancers and their biological processes, helping identify target biomarkers with prognostic and diagnostic values.

REFERENCES

1. S. Fernandez Tde, C. de Souza Fernandez and A. L. Mencialha, "Human induced pluripotent stem cells from basic research to potential clinical applications in cancer," *Biomed Res Int*, vol. 2013, pp. 430290, 2013.
2. D. Lai, Y. Wang, J. Sun, Y. Chen, T. Li, Y. Wu, L. Guo and C. Wei, "Derivation and characterization of human embryonic stem cells on human amnion epithelial cells," *Sci Rep*, vol. 5, pp. 10014, 2015.
3. M. Marti, L. Mulero, C. Pardo, C. Morera, M. Carrio, L. Laricchia-Robbio, C. R. Esteban and J. C. Izpisua Belmonte, "Characterization of pluripotent stem cells," *Nat Protoc*, vol. 8, no. 2, pp. 223-253, 2013.
4. M. Shelton, A. Kocharyan, J. Liu, I. S. Skerjanc and W. L. Stanford, "Robust generation and expansion of skeletal muscle progenitors and myocytes from human pluripotent stem cells," *Methods*, vol. 101, pp. 73-84, 2016.
5. A. Florkowska, I. Meszka, J. Nowacka, M. Granica, Z. Jablonska, M. Zawada, L. Truszkowski, M. A. Ciemerych and I. Grabowska, "PAX7 Balances the Cell Cycle Progression via Regulating Expression of Dnmt3b and Apobec2 in Differentiating PSCs," *Cells*, vol. 10, no. 9, 2021.
6. F. Relaix, D. Montarras, S. Zaffran, B. Gayraud-Morel, D. Rocancourt, S. Tajbakhsh, A. Mansouri, A. Cumano and M. Buckingham, "Pax3 and Pax7 have distinct and overlapping functions in adult muscle progenitor cells," *J Cell Biol*, vol. 172, no. 1, pp. 91-102, 2006.
7. C. A. Collins, V. F. Gnocchi, R. B. White, L. Boldrin, A. Perez-Ruiz, F. Relaix, J. E. Morgan and P. S. Zammit, "Integrated functions of Pax3 and Pax7 in the regulation of proliferation, cell size and myogenic differentiation," *PLoS One*, vol. 4, no. 2, pp. e4475, 2009.
8. N. Jiwlawat, E. Lynch, J. Jeffrey, J. M. Van Dyke and M. Suzuki, "Current Progress and Challenges for Skeletal Muscle Differentiation from Human Pluripotent Stem Cells Using Transgene-Free Approaches," *Stem Cells Int*, vol. 2018, pp. 6241681, 2018.
9. M. Schmidt, S. C. Schuler, S. S. Huttner, B. von Eyss and J. von Maltzahn, "Adult stem cells at work: regenerating skeletal muscle," *Cell Mol Life Sci*, vol. 76, no. 13, pp. 2559-2570, 2019.
10. Y. N. Jang and E. J. Baik, "JAK-STAT pathway and myogenic differentiation," *JAKSTAT*, vol. 2, no. 2, pp. e23282, 2013.
11. K. Ojima, "Myosin: Formation and maintenance of thick filaments," *Anim Sci J*, vol. 90, no. 7, pp. 801-807, 2019.
12. T. Hosoyama, J. Van Dyke and M. Suzuki, "Applications of skeletal muscle progenitor cells for neuromuscular diseases," *Am J Stem Cells*, vol. 1, no. 3, pp. 253-263, 2012.
13. Y. Kodaka, G. Rabu and A. Asakura, "Skeletal Muscle Cell Induction from Pluripotent Stem Cells," *Stem Cells Int*, vol. 2017, pp. 1376151, 2017.
14. J. C. Brown and L. A. Gilmore, "Physical Activity Reduces the Risk of Recurrence and Mortality in Cancer Patients," *Exerc Sport Sci Rev*, vol. 48, no. 2, pp. 67-73, 2020.
15. P. Hojman, C. Dethlefsen, C. Brandt, J. Hansen, L. Pedersen and B. K. Pedersen, "Exercise-induced muscle-derived cytokines inhibit mammary cancer cell growth," *Am J Physiol Endocrinol Metab*, vol. 301, no. 3, pp. E504-510, 2011.
16. P. Roy, S. Chowdhury and H. K. Roy, "Exercise-induced myokines as emerging therapeutic agents in colorectal cancer prevention and treatment," *Future Oncol*, vol. 14, no. 4, pp. 309-312, 2018.
17. P. M. Aponte and A. Caicedo, "Stemness in Cancer: Stem Cells, Cancer Stem Cells, and Their Microenvironment," *Stem Cells Int*, vol. 2017, pp. 5619472, 2017.
18. M. C. K. Severinsen and B. K. Pedersen, "Muscle-Organ Crosstalk: The Emerging Roles of Myokines," *Endocr Rev*, vol. 41, no. 4, 2020.
19. C. M. Dieli-Conwright, J. H. Parmentier, N. Sami, K. Lee, D. Spicer, W. J. Mack, F. Sattler and S. D. Mittelman, "Adipose tissue inflammation in breast cancer survivors: effects of a 16-week combined aerobic and resistance exercise training intervention," *Breast Cancer Res Treat*, vol. 168, no. 1, pp. 147-157, 2018.

20. K. Karstoft and B. K. Pedersen, "Skeletal muscle as a gene regulatory endocrine organ," *Curr Opin Clin Nutr Metab Care*, vol. 19, no. 4, pp. 270-275, 2016.
21. R. Schwappacher, W. Dieterich, D. Reljic, C. Pilarsky, D. Mukhopadhyay, D. K. Chang, A. V. Biankin, J. Siebler, H. J. Herrmann, M. F. Neurath and Y. Zopf, "Muscle-Derived Cytokines Reduce Growth, Viability and Migratory Activity of Pancreatic Cancer Cells," *Cancers (Basel)*, vol. 13, no. 15, 2021.
22. R. Schwappacher, K. Schink, S. Sologub, W. Dieterich, D. Reljic, O. Friedrich, H. J. Herrmann, M. F. Neurath and Y. Zopf, "Physical activity and advanced cancer: evidence of exercise-sensitive genes regulating prostate cancer cell proliferation and apoptosis," *J Physiol*, vol. 598, no. 18, pp. 3871-3889, 2020.
23. J. H. Lee and H. S. Jun, "Role of Myokines in Regulating Skeletal Muscle Mass and Function," *Front Physiol*, vol. 10, pp. 42, 2019.
24. M. A. Febbraio, N. Hiscock, M. Sacchetti, C. P. Fischer and B. K. Pedersen, "Interleukin-6 is a novel factor mediating glucose homeostasis during skeletal muscle contraction," *Diabetes*, vol. 53, no. 7, pp. 1643-1648, 2004.
25. A. Steensberg, C. Keller, R. L. Starkie, T. Osada, M. A. Febbraio and B. K. Pedersen, "IL-6 and TNF-alpha expression in, and release from, contracting human skeletal muscle," *Am J Physiol Endocrinol Metab*, vol. 283, no. 6, pp. E1272-1278, 2002.
26. S. Darkwah, E. J. Park, P. K. Myint, A. Ito, M. G. Appiah, G. Obeng, E. Kawamoto and M. Shimaoka, "Potential Roles of Muscle-Derived Extracellular Vesicles in Remodeling Cellular Microenvironment: Proposed Implications of the Exercise-Induced Myokine, Irisin," *Front Cell Dev Biol*, vol. 9, pp. 634853, 2021.
27. A. Festa, R. D'Agostino, Jr., G. Howard, L. Mykkanen, R. P. Tracy and S. M. Haffner, "Chronic subclinical inflammation as part of the insulin resistance syndrome: the Insulin Resistance Atherosclerosis Study (IRAS)," *Circulation*, vol. 102, no. 1, pp. 42-47, 2000.
28. M. A. Febbraio and B. K. Pedersen, "Contraction-induced myokine production and release: is skeletal muscle an endocrine organ?," *Exerc Sport Sci Rev*, vol. 33, no. 3, pp. 114-119, 2005.
29. S. Haberecht-Muller, E. Kruger and J. Fielitz, "Out of Control: The Role of the Ubiquitin Proteasome System in Skeletal Muscle during Inflammation," *Biomolecules*, vol. 11, no. 9, 2021.
30. W. R. Frontera and J. Ochala, "Skeletal muscle: a brief review of structure and function," *Calcif Tissue Int*, vol. 96, no. 3, pp. 183-195, 2015.
31. M. A. Egerman and D. J. Glass, "Signaling pathways controlling skeletal muscle mass," *Crit Rev Biochem Mol Biol*, vol. 49, no. 1, pp. 59-68, 2014.
32. D. Peris-Moreno, L. Cussonneau, L. Combaret, C. Polge and D. Taillandier, "Ubiquitin Ligases at the Heart of Skeletal Muscle Atrophy Control," *Molecules*, vol. 26, no. 2, 2021.
33. E. R. Sanchez, R. Sanchez and C. Moliver, "Anatomic relationship of the pectoralis major and minor muscles: a cadaveric study," *Aesthet Surg J*, vol. 34, no. 2, pp. 258-263, 2014.
34. N. Samreen, C. Lee, A. Bhatt, J. Carter, T. Hieken, K. Adler, S. Zingula and K. N. Glazebrook, "A Clinical Approach to Diffusion-Weighted Magnetic Resonance Imaging in Evaluating Chest Wall Invasion of Breast Tumors," *J Clin Imaging Sci*, vol. 9, pp. 11, 2019.
35. J. Bohlen, S. L. McLaughlin, H. Hazard-Jenkins, A. M. Infante, C. Montgomery, M. Davis and E. E. Pistilli, "Dysregulation of metabolic-associated pathways in muscle of breast cancer patients: preclinical evaluation of interleukin-15 targeting fatigue," *J Cachexia Sarcopenia Muscle*, vol. 9, no. 4, pp. 701-714, 2018.
36. H. Dastsooz, M. Cereda, D. Donna and S. Oliviero, "A Comprehensive Bioinformatics Analysis of UBE2C in Cancers," *Int J Mol Sci*, vol. 20, no. 9, 2019.
37. K. Ieta, E. Ojima, F. Tanaka, Y. Nakamura, N. Haraguchi, K. Mimori, H. Inoue, H. Kuwano and M. Mori, "Identification of overexpressed genes in hepatocellular carcinoma, with special reference to ubiquitin-conjugating enzyme E2C gene expression," *Int J Cancer*, vol. 121, no. 1, pp. 33-38, 2007.

38. P. Pallante, M. T. Berlingieri, G. Troncione, M. Kruhoffer, T. F. Orntoft, G. Viglietto, A. Caleo, I. Migliaccio, M. Decaussin-Petrucci, M. Santoro, L. Palombini and A. Fusco, "UbcH10 overexpression may represent a marker of anaplastic thyroid carcinomas," *Br J Cancer*, vol. 93, no. 4, pp. 464-471, 2005.
39. T. Fujita, H. Ikeda, N. Taira, S. Hatoh, M. Naito and H. Doihara, "Overexpression of UbcH10 alternates the cell cycle profile and accelerates the tumor proliferation in colon cancer," *BMC Cancer*, vol. 9, pp. 87, 2009.
40. M. T. Berlingieri, P. Pallante, A. Sboner, M. Barbareschi, M. Bianco, A. Ferraro, G. Mansueto, E. Borbone, E. Guerriero, G. Troncione and A. Fusco, "UbcH10 is overexpressed in malignant breast carcinomas," *Eur J Cancer*, vol. 43, no. 18, pp. 2729-2735, 2007.
41. I. Perrotta, L. Bruno, L. Maltese, E. Russo, A. Donato and G. Donato, "Immunohistochemical analysis of the ubiquitin-conjugating enzyme UbcH10 in lung cancer: a useful tool for diagnosis and therapy," *J Histochem Cytochem*, vol. 60, no. 5, pp. 359-365, 2012.
42. L. Jiang, C. G. Huang, Y. C. Lu, C. Luo, G. H. Hu, H. M. Liu, J. X. Chen and H. X. Han, "Expression of ubiquitin-conjugating enzyme E2C/UbcH10 in astrocytic tumors," *Brain Res*, vol. 1201, pp. 161-166, 2008.
43. M. V. Bose, G. Gopisetty, G. Selvaluxmy and T. Rajkumar, "Dominant negative Ubiquitin-conjugating enzyme E2C sensitizes cervical cancer cells to radiation," *Int J Radiat Biol*, vol. 88, no. 9, pp. 629-634, 2012.
44. M. Akutsu, I. Dikic and A. Bremm, "Ubiquitin chain diversity at a glance," *J Cell Sci*, vol. 129, no. 5, pp. 875-880, 2016.
45. M. J. Clague and S. Urbe, "Ubiquitin: same molecule, different degradation pathways," *Cell*, vol. 143, no. 5, pp. 682-685, 2010.
46. H. J. Meyer and M. Rape, "Enhanced protein degradation by branched ubiquitin chains," *Cell*, vol. 157, no. 4, pp. 910-921, 2014.
47. M. J. Clague, I. Barsukov, J. M. Coulson, H. Liu, D. J. Rigden and S. Urbe, "Deubiquitylases from genes to organism," *Physiol Rev*, vol. 93, no. 3, pp. 1289-1315, 2013.
48. M. J. Clague, C. Heride and S. Urbe, "The demographics of the ubiquitin system," *Trends Cell Biol*, vol. 25, no. 7, pp. 417-426, 2015.
49. A. Rousseau and A. Bertolotti, "Regulation of proteasome assembly and activity in health and disease," *Nat Rev Mol Cell Biol*, vol. 19, no. 11, pp. 697-712, 2018.
50. M. Schmidt and D. Finley, "Regulation of proteasome activity in health and disease," *Biochim Biophys Acta*, vol. 1843, no. 1, pp. 13-25, 2014.
51. J. Beers, D. R. Gulbranson, N. George, L. I. Siniscalchi, J. Jones, J. A. Thomson and G. Chen, "Passaging and colony expansion of human pluripotent stem cells by enzyme-free dissociation in chemically defined culture conditions," *Nat Protoc*, vol. 7, no. 11, pp. 2029-2040, 2012.
52. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson and T. R. Gingeras, "STAR: ultrafast universal RNA-seq aligner," *Bioinformatics*, vol. 29, no. 1, pp. 15-21, 2013.
53. Y. Liao, G. K. Smyth and W. Shi, "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features," *Bioinformatics*, vol. 30, no. 7, pp. 923-930, 2014.
54. M. D. Robinson, D. J. McCarthy and G. K. Smyth, "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data," *Bioinformatics*, vol. 26, no. 1, pp. 139-140, 2010.
55. Z. Gu, R. Eils and M. Schlesner, "Complex heatmaps reveal patterns and correlations in multidimensional genomic data," *Bioinformatics*, vol. 32, no. 18, pp. 2847-2849, 2016.
56. G. Yu, L. G. Wang, Y. Han and Q. Y. He, "clusterProfiler: an R package for comparing biological themes among gene clusters," *OMICS*, vol. 16, no. 5, pp. 284-287, 2012.
57. H. E. Wilson, D. A. Stanton, S. Rellick, W. Geldenhuys and E. E. Pistilli, "Breast cancer-associated skeletal muscle mitochondrial dysfunction and lipid accumulation is reversed by PPAR γ ," *Am J Physiol Cell Physiol*, vol. 320, no. 4, pp. C577-C590, 2021.

58. H. Dastsooz, A. Alizadeh, P. Habibzadeh, A. Nariman, A. Hosseini, Y. Mansoori and H. Haghi-Aminjan, "LncRNA-miRNA-mRNA Networks of Gastrointestinal Cancers Representing Common and Specific LncRNAs and mRNAs," *Front Genet*, vol. 12, pp. 791919, 2021.
59. R. Li, H. Qu, S. Wang, J. Wei, L. Zhang, R. Ma, J. Lu, J. Zhu, W. D. Zhong and Z. Jia, "GDCRNATools: an R/Bioconductor package for integrative analysis of lncRNA, miRNA and mRNA data in GDC," *Bioinformatics*, vol. 34, no. 14, pp. 2515-2517, 2018.
60. E. Y. Chen, C. M. Tan, Y. Kou, Q. Duan, Z. Wang, G. V. Meirelles, N. R. Clark and A. Ma'ayan, "Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool," *BMC Bioinformatics*, vol. 14, pp. 128, 2013.
61. M. D. Stewart, T. Ritterhoff, R. E. Klevit and P. S. Brzovic, "E2 enzymes: more than just middle men," *Cell Res*, vol. 26, no. 4, pp. 423-440, 2016.
62. E. Bober, T. Franz, H. H. Arnold, P. Gruss and P. Tremblay, "Pax-3 is required for the development of limb muscles: a possible role for the migration of dermomyotomal muscle progenitor cells," *Development*, vol. 120, no. 3, pp. 603-612, 1994.
63. B. A. Williams and C. P. Ordahl, "Pax-3 expression in segmental mesoderm marks early stages in myogenic cell specification," *Development*, vol. 120, no. 4, pp. 785-796, 1994.
64. H. B. Wood and V. Episkopou, "Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages," *Mech Dev*, vol. 86, no. 1-2, pp. 197-201, 1999.
65. V. Vilmont, B. Cadot, G. Ouanounou and E. R. Gomes, "A system for studying mechanisms of neuromuscular junction development and maintenance," *Development*, vol. 143, no. 13, pp. 2464-2477, 2016.
66. L. L. Arnold, A. Cecchini, D. A. Stark, J. Ihnat, R. N. Craigg, A. Carter, S. Zino and D. Cornelison, "EphA7 promotes myogenic differentiation via cell-cell contact," *Elife*, vol. 9, 2020.
67. M. A. Farazi Fard, A. P. Rebelo, E. Buglo, H. Nemat, H. Dastsooz, I. Gehweiler, S. Reich, J. Reichbauer, B. Quintans, A. Ordóñez-Ugalde, A. Cortese, S. Courel, L. Abreu, E. Powell, M. C. Danzi, N. B. Martuscelli, D. M. Bis-Brewer, F. Tao, F. Zarei, P. Habibzadeh, M. Yavarian, F. Modarresi, M. Silawi, Z. Tabatabaei, M. Yousefi, H. R. Farpour, C. Kessler, E. Mangold, X. Kobeleva, I. Tournev, T. Chamova, A. J. Mueller, T. B. Haack, M. Tarnopolsky, Z. Gan-Or, G. A. Rouleau, M. Synofzik, M. J. Sobrido, A. Jordanova, R. Schule, S. Zuchner and M. A. Faghihi, "Truncating Mutations in UBAP1 Cause Hereditary Spastic Paraplegia," *Am J Hum Genet*, vol. 104, no. 6, pp. 1251, 2019.
68. D. Barets, R. Appay, M. Heinisch, M. Battistella, C. Bouvier, G. Chotard, F. Le Loarer, N. Macagno, R. Perbet, D. Pissaloux, A. Rousseau, A. Tauziède-Espariat, P. Varlet, A. Vasiljevic, C. Colin, F. Fina and D. Figarella-Branger, "Specific and Sensitive Diagnosis of BCOR-ITD in Various Cancers by Digital PCR," *Front Oncol*, vol. 11, pp. 645512, 2021.
69. S. Jain and A. Abraham, "BCR-ABL1-like B-Acute Lymphoblastic Leukemia/Lymphoma: A Comprehensive Review," *Arch Pathol Lab Med*, vol. 144, no. 2, pp. 150-155, 2020.
70. Y. Hu, Z. Song, J. Chen and C. Caulin, "Overexpression of MYB in the Skin Induces Alopecia and Epidermal Hyperplasia," *J Invest Dermatol*, vol. 140, no. 6, pp. 1204-1213 e1205, 2020.
71. J. Lin, S. Fu, C. Yang and C. Redies, "Pax3 overexpression induces cell aggregation and perturbs commissural axon projection during embryonic spinal cord development," *J Comp Neurol*, vol. 525, no. 7, pp. 1618-1632, 2017.
72. L. Santiago, G. Daniels, D. Wang, F. M. Deng and P. Lee, "Wnt signaling pathway protein LEF1 in cancer, as a biomarker for prognosis and a target for treatment," *Am J Cancer Res*, vol. 7, no. 6, pp. 1389-1406, 2017.
73. D. He, X. Zhang and J. Tu, "Diagnostic significance and carcinogenic mechanism of pan-cancer gene POU5F1 in liver hepatocellular carcinoma," *Cancer Med*, vol. 9, no. 23, pp. 8782-8800, 2020.
74. A. Calzolari, V. Finisguerra, I. Oliviero, S. Deaglio, G. Mariani, F. Malavasi and U. Testa, "Regulation of transferrin receptor 2 in human cancer cell lines," *Blood Cells Mol Dis*, vol. 42, no. 1, pp. 5-13, 2009.

75. W. J. Tissing, M. L. den Boer, J. P. Meijerink, R. X. Menezes, S. Swagemakers, P. J. van der Spek, S. E. Sallan, S. A. Armstrong and R. Pieters, "Genomewide identification of prednisolone-responsive genes in acute lymphoblastic leukemia cells," *Blood*, vol. 109, no. 9, pp. 3929-3935, 2007.

Part two

PUBLICATIONS (Ph.D. program, 2018-2021)

Abstract

During my PhD program I also studied some bioinformatic analyses for human genetic disorders, including human cancers and neuromuscular disorders. For example, I used different bioinformatic tools and R packages to analyze expression of mRNA, lncRNAs, and miRNA and their networks involved in cancers and could publish interesting data related to *UBE2C* and other genes in cancers. Moreover, I had some collaborations with other researchers to analyze the annotated data of whole exome sequencing (WES) and disorder panel sequencing of human genetic disorders, design some studies, and write the manuscripts. In these NGS data, I could find, correlate and report some mutated genes (such as *FUT8*, *FANCF*, and *SEPNI*, among others) related to rare genetic metabolic, neuromuscular, and neurologic disorders and strengthen their contributions to these disorders. In the most of these published studies, I had substantial contributions for the NGS data analysis, bioinformatic analyses related to the confirmation of mutations, sanger sequencing analysis, protein network analysis, conceptualization of some studies, and writing and editing the manuscripts.



HHS Public Access

Author manuscript

J Inherit Metab Dis. Author manuscript; available in PMC 2020 July 14.

Published in final edited form as:

J Inherit Metab Dis. 2020 July ; 43(4): 871–879. doi:10.1002/jimd.12221.

Expanding the molecular and clinical phenotypes of FUT8-CDG

Bobby G. Ng^{#1}, Hassan Dastsooz^{#2,3}, Mohammad Silawi³, Parham Habibzadeh³, Shima B. Jahan³, Mohammad A. F. Fard³, Benjamin J. Halliday⁴, Kimiyo Raymond⁵, Maura R. Z. Ruzhnikov^{6,7}, Zahra Tabatabaei³, Afsaneh Taghipour-Sheshdeh³, Elise Brimble⁶, Stephen P. Robertson⁴, Mohammad A. Faghihi^{3,8}, Hudson H. Freeze¹

¹Human Genetics Program, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, California ²Department of Life Sciences and Systems Biology, Italian Institute for Genomic Medicine (IIGM), University of Turin, Turin, Italy ³Persian BayanGene Research and Training Center, Shiraz University of Medical Sciences, Shiraz, Iran ⁴Department of Women's and Children's Health, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand ⁵Biochemical Genetics Laboratory, Mayo Clinic College of Medicine, Rochester, Minnesota ⁶Department of Neurology and Neurological Sciences, Stanford Medicine, Stanford, California ⁷Division of Medical Genetics, Department of Pediatrics, Stanford Medicine, Stanford, California ⁸Center for Therapeutic Innovation, Department of Psychiatry and Behavioral Sciences, University of Miami Miller School of Medicine, Miami, Florida

These authors contributed equally to this work.

Abstract

Pathogenic variants in the Golgi localised alpha 1,6 fucosyltransferase, *FUT8*, cause a rare inherited metabolic disorder known as FUT8-CDG. To date, only three affected individuals have been reported presenting with a constellation of symptoms including intrauterine growth restriction, severe delays in growth and development, other neurological impairments, significantly shortened limbs, respiratory complications, and shortened lifespan. Here, we report an additional four unrelated affected individuals homozygous for novel pathogenic variants in *FUT8*. Analysis

Correspondence: Hudson H. Freeze, Human Genetics Program, Sanford Children's Health Research Center, Sanford-Burnham-Prebys Medical Discovery Institute, 10901 N. Torrey Pines Rd. La Jolla, CA 92037. hudson@sbsdsc.org.

ROLE FOR EACH CONTRIBUTING AUTHOR

B.N. performed experiments and drafted manuscript. H.D., M.S., and M.A.F. performed NGS data analysis of CDG-0095 and drafted manuscript. H.D., M.S., P.H., S.B.J., M.A.F.F., M.R., E.B., Z.T., S.R., A.T.S., M.A.F. provided clinical evaluations and drafted manuscript. B.H. provides bioinformatic analysis of whole-genomes sequence data for CDG-0099. K.R. performed MALDI-TOF/TOF-MS analysis and interpretation. H.F. supervised and drafted manuscript.

CONFLICT OF INTEREST

Bobby G. Ng, Hassan Dastsooz, Mohammad Silawi, Parham Habibzadeh, Shima Bahram Jahan, Mohammad Ali Farazi Fard, Benjamin Halliday, Kimiyo Raymond, Maura R. Z. Ruzhnikov, Zahra Tabatabaei, Afsaneh Taghipour-Sheshdeh, Elise Brimble, Stephen P. Robertson, Mohammad Ali Faghihi and Hudson H. Freeze declare that they have no conflict of interest.

ETHICAL APPROVAL AND INFORMED CONSENT

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000 (5). Informed consent was obtained from all patients for being included in the study. Sanford Burnham Prebys Medical Discovery Institute (IRB-2014-038-17).

ANIMAL RIGHTS

This article does not contain any studies with animal subjects.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

of serum *N*-glycans revealed a complete lack of core fucosylation, an important diagnostic biomarker of FUT8-CDG. Our data expands both the molecular and clinical phenotypes of FUT8-CDG and highlights the importance of identifying a reliable biomarker for confirming potentially pathogenic variants.

Keywords

congenital disorders of glycosylation; core fucosylation; mass spectrometry; *N*-glycans; whole exome sequencing

1 | INTRODUCTION

Congenital disorders of glycosylation (CDG) are a clinically diverse group of primarily autosomal recessive disorders that result in altered protein and lipid glycosylation. To date, more than 135 genes have been found to cause CDG.^{1,2} Several of these disorders are specifically due to abnormalities in fucosylation, which is an enzymatic process of incorporating the monosaccharide L-fucose into the *N*- and *O*-glycans of proteins or glycolipids.³

Proper fucosylation requires the formation of the activated donor substrate, guanosine diphosphate L-fucose (GDP-fucose), which can occur by both *de novo* and salvage mechanisms.³ It is widely believed that the *de novo* pathway accounts for the vast majority (90%) of GDP-fucose generated within a cell.⁴ In the *de novo* pathway, glucose and mannose are essential for synthesising GDP-mannose, of which a portion is subsequently converted to GDP-fucose via a two-enzyme mechanism involving GDP mannose 4,6-dehydratase (GMD5), and GDP-keto-6-deoxymannose 3,5 epimerase (TSTA3) (aka FX protein).³ The salvage pathway also utilises a two-enzyme mechanism to convert exogenously provided (diet) or recycled L-fucose (lysosomal-dependent glycan degradation) to GDP-fucose. Here, L-fucose is first phosphorylated by fucose kinase (FCSK) to generate fucose-1-phosphate which then undergoes a pyrophosphorylase (FPGT) reaction to yield GDP-fucose.^{3,5} Postnatal lethality in a *Tsta3* (aka FX) knockout (KO) mouse model has shown that complete loss of the *de novo* pathway can be rescued by providing exogenous L-fucose, suggesting a compensatory role of the salvage pathway.⁶ Ultimately, GDP-fucose is transported into the Golgi via SLC35C1 or into the endoplasmic reticulum by SLC35C2 where specific fucosyltransferases add fucose to various acceptor substrates.^{7,8}

The most ubiquitous form of fucosylation is “Core fucosylation” of *N*-linked glycans and requires the Golgi localised fucosyltransferase FUT8 to attach an L-fucose to the *N*-acetylglucosamine (GlcNAc) directly linked to an asparagine (Asn)³ (Figure 1A). FUT8 contains three distinct domains, an N-terminal coiled-coil domain, a glycosyltransferase 23 family domain and a C-terminal SH3 domain. It is unclear what role the coiled-coil or SH3 domains of FUT8 play in its function. Although, if compared to other proteins containing these domains, they likely are involved in protein-protein interactions.¹⁰ Core fucosylation has been shown to be critical for many cellular processes including immune system regulation, inflammatory responses, cancer metastasis, and embryonic development.^{11–15}

Thus far, biochemically confirmed pathogenic variants have been identified in fucose kinase (*FCSK*) [OMIM #618324], fucosyltransferase 8 (*FUT8*) [OMIM# 618005], GDP-fucose transporter (*SLC35C1*) [OMIM# 605881], protein O-fucosyltransferases (*POFUT1*) [OMIM# 615327], fucose-specific beta-1,3-N-acetylglucosaminyltransferase (*LFNG*) [OMIM# 609813], and the lysosomal fucosidase 1 (*FUCA1*) [OMIM# 612280]. All are extremely rare disorders, and only a few *SLC35C1*-CDG individuals have been shown to benefit from oral fucose therapy.¹⁶

Here, we identify four unreported *FUT8*-CDG subjects expanding this disorder's molecular and clinical presentation.

2 | METHODS AND MATERIALS

2.1 | Clinical data

Written consent was provided for all four families in accordance with a Sanford Burnham Prebys Medical Discovery Institute approved IRB-2014-038-17.

2.2 | Serum N-glycan analysis

Matrix-assisted laser desorption/ionisation tandem time of flight/time of flight (MALDI-TOF/TOF) mass spectrometry analysis of total serum N-glycans was performed as previously described at the Mayo Clinic.¹⁷

2.3 | Exome sequencing and analysis

All four affected individuals had next generation sequencing performed by different institutions. CDG-0099 had whole genome trio sequencing performed in a research setting, while CDG-0108 had whole exome sequencing (WES) performed by a CLIA approved clinical lab service. Specifically, for CDG-0095 and 0096, genomic DNA was extracted from blood and used for WES on an Illumina NextSeq500 instrument with 150-nucleotide paired-end sequencing. WES raw data were processed and analysed with BWA aligner (19451168), GATK (20644199), and ANNOVAR (20601685). All variants were confirmed by Sanger sequencing.

3 | RESULTS

3.1 | Clinical phenotype

CDG-0095 is a female infant of Iranian ancestry born to consanguineous second-degree cousins (Figure 1B). The family history was notable for an affected sibling with similar presentation who passed away at 14 months of age without a diagnosis. The pregnancy was complicated by polyhydramnios. However, there was no evidence of intrauterine growth restriction (IUGR) on prenatal ultrasonography. She was unable to tolerate breastfeeding and had hypotonia and dysmorphic features on physical examination (Table 1). She had global developmental delays with cognitive impairment and epilepsy. She had a history of recurrent pneumonia. She is still alive at 3 years of age.

The second individual, CDG-0096, is a male also of Iranian descent and a product of a second-degree consanguineous marriage (Figure 1B). The pregnancy was not complicated by polyhydramnios or IUGR, but he was delivered at 37 weeks gestation by caesarean section due to meconium aspiration. During the postnatal period, he was hypotonic and could not tolerate breast feeding. He was also found to have dysplastic ears and dysmorphic facial features. He has a history of developmental delays with severe intellectual disability and epilepsy that is only partially controlled with medication (Table 1). Brain magnetic resonance imaging revealed slight ventriculomegaly with mildly delayed myelination. Abdominopelvic ultrasonography showed renal stones in both kidneys without any evidence of hepatosplenomegaly. The individual has had recurrent pneumonia for which he was admitted to hospital on two occasions. Thyroid-stimulating hormone level ($8.1 \mu\text{U/mL}$, reference range: $0.3\text{--}5.5 \mu\text{U/mL}$) was mildly increased. He remains alive at 6 years of age.

The third individual, CDG-0099, is an affected female from healthy non-consanguineous parents of European ancestry who did not have a family history of a similar disorder (Figure 1B). No pregnancy complications were reported; however, IUGR was noted. She presented early with significant hypotonia, facial dysmorphism, failure to thrive, and feeding problems requiring placement of a nasogastric tube (NG-tube). She has a history of severe developmental delays with intellectual disability, along with intractable seizures. Skeletal abnormalities include short stature and scoliosis. She is noted to have congenital glaucoma (Table 1). CDG-0099 is still alive at 4 years of age.

Finally, CDG-0108 is an affected female born to healthy consanguineous parents of Southeast Asian ancestry (Figure 1B). No complications or IUGR were noted during pregnancy. She has significant hypotonia and facial dysmorphism that includes a flattened nasal bridge with short rounded tip and broad lips with a lower lip that is everted. Like the previously mentioned subjects, she has feeding problems with severe failure to thrive, a history of global developmental delay with severe intellectual disabilities, epilepsy, and microcephaly. A multitude of skeletal abnormalities were noted including contractures of the fingers, ankles and feet, progressive scoliosis and a history of left hip dislocations (Table 1). Like those described above, CDG-0108 is still alive at 16 years of age, making her the oldest known FUT8-CDG case.

Comparing the three previously reported FUT8-CDG cases, these four reveal striking similarities. All seven individuals presented with severe global developmental delays with cognitive impairment or intellectual disability, feeding problems, dysmorphic facial features, microcephaly, seizures, hypotonia, and various skeletal abnormalities (Figure 2, Table 1). This suggests that FUT8-CDG could have a recognisable phenotypic presentation. Respiratory difficulties were also seen in 7/7 individuals with five of those seven having recurrent infections (Figure 2, Table 1). This is noteworthy because FUT8-KO mice also have respiratory difficulties.¹⁵ Cardiac and renal abnormalities were less frequent and could be due to other genetic factors. While all three original cases presented with IUGR and ultimately died prematurely, only one of the current four had IUGR and all are still alive (Figure 2, Table 1).

In all four cases, clinical phenotyping and laboratory testing failed to reach a definitive diagnosis. Therefore, next-generation sequencing was used to reach a diagnosis.

3.2.1 Molecular analysis

All four affected individuals had next generation sequencing performed at different institutions or clinical lab services. Three of the four individuals (0095, 0096, 0108) were from consanguineous families who had whole exome performed on the proband (Table 1). For these individuals, due to the consanguinity, filtering was initially focused on rare homozygous variants of less than 1% allele frequency. CDG-0099 had whole genome sequencing performed on the family trio, which was used to determine potential compound heterozygous, homozygous, and de novo variants that segregated appropriately.

Importantly, variant filtering revealed that three of the four individuals (0095, 0099, 0108) were homozygous for truncating loss of function INDEL or nonsense variants in the alpha 1,6 fucosyltransferase *FUT8* (NCBI Refseq—NM_178155.2, Uniprot—Q9BYC5). CDG-0095 was identified to have a novel c.1302G > A [p.Trp434*], CDG-0099 a novel c.1402delT [p.Ser468Hisfs*26] and CDG-0108 c.1675C > T [p.Arg559*] (Table 1, Figure 1B,C). Interestingly, CDG-0096 was determined to be homozygous for a variant of uncertain significance (VUS) c.716G > A [p.R239Q], which overlaps with a previously reported nonsense variant c.715C > T (p.Arg239*)⁹ (Table 1, Figure 1B, C). Three in silico prediction programs suggest the c.716G > A [p.R239Q] to be damaging with a Polyphen2 (<http://genetics.bwh.harvard.edu/pph2/>) score of 1 (probably damaging), SIFT (<http://provean.jcvi.org/index.php>) score of 0 (deleterious), and a CADD (<https://cadd.gs.washington.edu/>) score of 26.1 which would place it in the top 0.5% of deleterious variants in the human genome.

Only the c.1675C > T [p.Arg559*] variant was seen in a single heterozygous carrier from the gnomAD database (1/250 464 alleles) (<http://gnomad.broadinstitute.org/>) (gnomAD v2.1.1 accessed November 25, 2019) of 125 748 exomes and 15 708 whole-genomes of unrelated individuals. Neither of the four variants were seen in gnomAD v3, which is composed of 71 702 genomes (doi: <https://doi.org/10.1101/531210>).

3.3.1 Serum N-glycan analysis

We had previously used MALDI-TOF/TOF on serum samples to confirm the lack of core fucosylation in individuals with *FUT8*-CDG.⁹ We applied this same technique to three of the four affected individuals presented here. The lone individual (CDG-0099) we were unable to obtain sample for and thus did not have MALDI-TOF/TOF-MS of serum glycoproteins was homozygous for the c.1402delT [p.Ser468Hisfs*26] (Table 1). We have previously shown that truncating variants in *FUT8*, like that seen in CDG-0108, are not tolerated and results in the complete loss of core fucosylation, ultimately causing *FUT8*-CDG.⁹ When compared to a representative control serum sample from a “healthy” individual, CDG-0095 and CDG-0108 who are both homozygous for nonsense mutations, showed a complete loss of core fucosylated *N*-glycans from serum proteins (Figure 3). MALDI-TOF/TOF-MS values highlighted within red circles represent those glycans that are completely missing in both CDG-0095 and CDG-0108 (1836; 2040; 2081, 2244, 2606; 2967 m/z) (Figure 3).

Importantly, CDG-0096, who carries the VUS c.716G > A [p.R239Q] was also shown to have a complete loss of core fucosylation (Figure 3). Thus, confirming that the c.716G > A [p.R239Q] is in fact, a pathogenic variant. Finally, analysis of serum transferrin glycosylation in at least one individual was normal and did not reveal any clear abnormalities in fucosylation (Figure S1). However, it should be noted that only a small percentage of the total transferrin undergoes fucosylation.

4 | DISCUSSION

There are an estimated 13 fucosyltransferases encoded in the human genome, but only one (*FUT8*) encodes an enzyme capable of carrying out core fucosylation.³ The importance of creating the FUT8-dependent core fucosylation epitope can be seen in both mouse KO models and individuals who have FUT8-CDG. *Fut8* KO mice show severe growth restriction, respiratory defects, and a high mortality rate of ~70%.¹⁵ KO mice that do survive are more vulnerable to progressive chronic obstructive pulmonary disease and have been shown to exhibit a significant reduction in liver regeneration capabilities following partial hepatectomy.^{18,19} Additionally, they display increased neuroinflammation and display multiple behavioural abnormalities consistent with a schizophrenia-like phenotype.^{13,20} Consistent with the KO mouse model, affected individuals with FUT8-CDG show severe developmental and growth delays, shortened limbs, various neurological impairments, and respiratory complications.⁹ Subsequent follow up of the original reported cases revealed that all three have passed away, while all four subjects described here are still alive at 3 to 16 years of age.

Interestingly, when heterozygous *Fut8*^{+/-} mice were given L-fucose prior to partial hepatectomies, those that received L-Fucose saw dramatically accelerated liver regeneration.¹⁹ This likely occurred through increased salvage-dependent synthesis of GDP-fucose. We had previously tested the effects of supplementing tissue culture medium with L-fucose on FUT8-CDG fibroblasts, but we saw no increase in total fucosylation.⁹ We considered L-fucose supplementation with CDG-0096 since he had a missense variant that could potentially have some residual activity. However, total *N*-glycan analysis of serum showed that this individual was not capable of synthesising any detectable core fucosylation (Figure 3). Providing fucose could potentially create a neo-antigen specific for that individual and could precipitate a severe immunological response. The family and physicians decided the risks outweighed the potential benefits of fucose supplementation.

In summary, FUT8-CDG is an extremely rare metabolic disorder with only three documented cases who present with a multitude of symptoms including IUGR, severe delays in growth and development, severe neurological impairments, significantly shortened limbs, respiratory complications, and decreased lifespan. Here, we present on an additional four unrelated affected individuals to expand both the molecular and clinical knowledge for this disorder and highlight the usefulness of a reliable serum biomarker for confirming FUT8-CDG cases.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We would like to thank all the families for their continued support and for providing valuable biological specimens. This work is supported by The Rocket Fund and The Bella Conneran Foundation, National Institutes of Health (NIH) grants R01DK099551 (to H.H.F). SPR is supported by Curekids (NZ) and Genomics Aotearoa. This paper is dedicated to the memory of Bella Conneran.

Funding information

Genomics Aotearoa; Curekids; National Institutes of Health, Grant/Award Number: R01DK099551; Bella Conneran Foundation; Rocket Fund

Ng et al.

Page 9

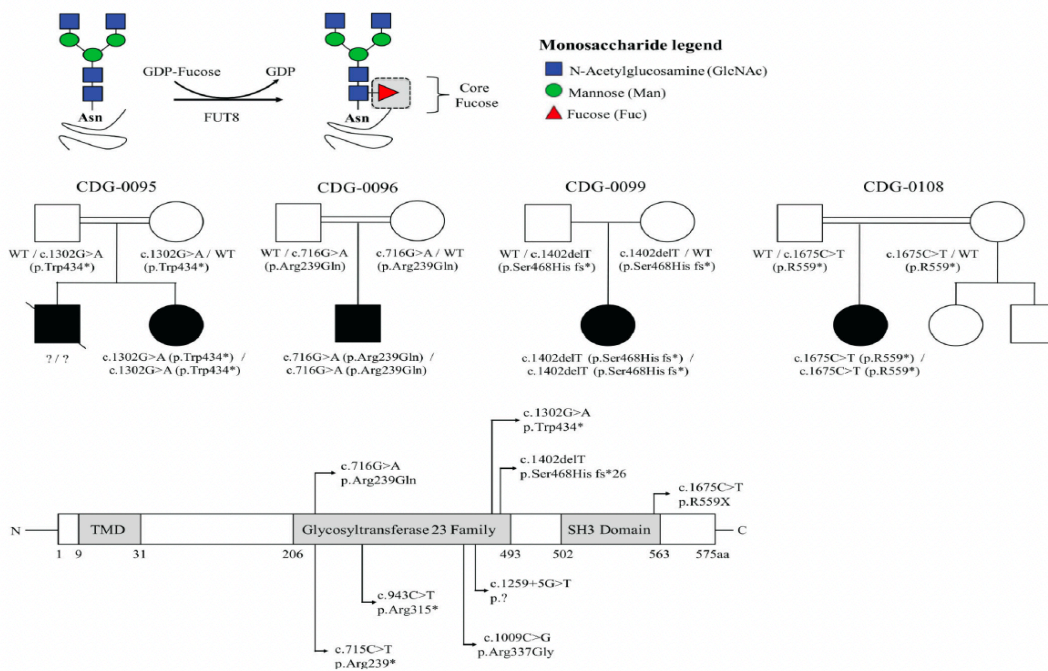


FIGURE 1. Identification of novel *FUT8* pathogenic variants in four unrelated families. A, Schematic showing the *FUT8*-dependent synthesis of core fucose. B, Pedigrees for the four families. C, Schematic showing both new (upper) and previously reported (lower) mutations in *FUT8* (GenBank: NM_178155.2, Uniprot: Q9BYC5). Figure adapted and updated from Reference 9

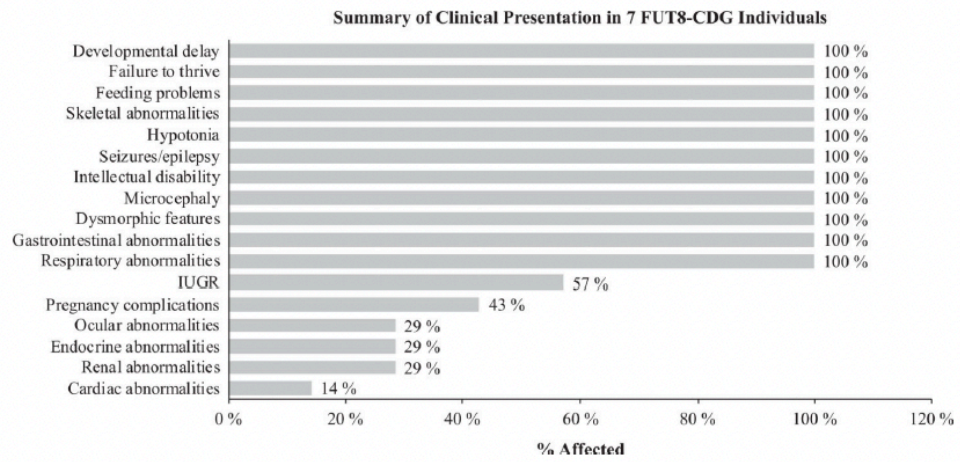


FIGURE 2. Clinical Summary for the three previously reported and four new FUT8-CDG individuals

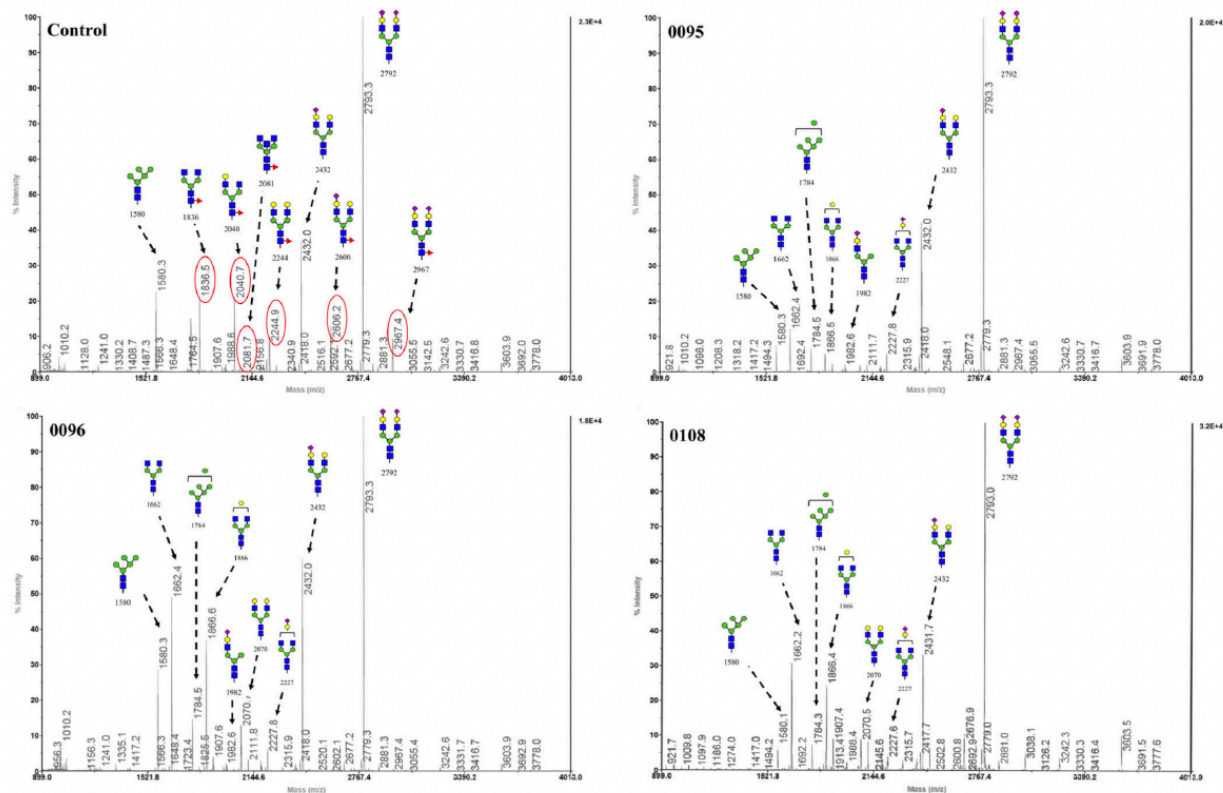


FIGURE 3. MALDI-TOF/TOF-MS *N*-glycan profiles from a healthy control serum sample as well as CDG-0095, 0096, 0108. Highlighted within the red circles of the control sample, are those core fucosylated *N*-glycans that are completely lost in each of the three (0095, 0096, 0108) affected FUT8-CDG individuals

TABLE 1
Clinical and molecular summaries for the three previously reported and four new FUT8-CDG individuals

	Ng et al ^p	Ng et al ^p	Ng et al ^p	Ng et al ^p	CDG—0095	CDG—0096	CDG—0099	CDG—0108	Total
Gender	Female	Female	Male	Female	Male	Female	Female	Female	5F/2M
Year of birth	2014	2014	2010	2016	2013	2015	2015	2003	
Death (date)	Feb 2018	Feb 2018	Sept 2017	NA-Alive	NA-Alive	NA-Alive	NA-Alive	NA-Alive	
Family history	No	No	Yes, deceased older brother	Older brother with similar presentation, deceased at 14 months	Intellectual disability in distant relatives	No	No	No	
Consanguinity	Yes, second cousin	No	Yes, first cousin	Yes, second cousin	Yes, second cousin	No	No	Yes, first cousins	
Ancestry	Irish, Italian, Native American	Irish-Norwegian / Irish-Ukrainian-Belgian	Arab	Iranian	Iranian	New Zealand / European	New Zealand / European	Southeast Asia (Philippines)	
Detection method	WES	WES	WES	WES	WES	WGS	WGS	WES	
FUT8 pathogenic variants	c.715C > T (p.Arg239*)	c.1009C > G (p.Arg337Gly) c.1259 + 5 G > T (N/A)	c.943C > T (p.Arg315*)	c.1302G > A (p.Trp434*)	c.716G > A (p.Arg239Gln)	c.1402delT (p.Ser468His fs*26)	c.1675C > T (p.R559*)		
Developmental delay	Severe	Severe	Severe	Severe	Severe	Severe	Severe	Severe	7/7
Failure to thrive	Severe	Severe	Severe	Severe	Severe	Severe	Severe	Severe	7/7
Feeding problems	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	7/7
Skeletal abnormalities	Short stature	Short stature	Short stature, multiple contractures, severe osteopenia, hip dislocations, kyphoscoliosis	Short stature, upper extremity hemihypertrophy	Short stature	Short stature, scoliosis	Short stature, scoliosis	Contractures fingers, ankles, feet; progressive scoliosis; left hip dislocation; short stature	7/7
Hypotonia	Yes	Yes	Central with peripherical spasticity	Yes	Yes	Yes	Yes	Yes	7/7
Seizures/epilepsy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	7/7
Intellectual disability	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	7/7
Microcephaly	Yes, 33 months <first percentile	Yes, 31 months <10th percentile	Yes, 36 months <third percentile	Yes, 21 months <first percentile	Yes, 5 Years <third percentile	Yes	Yes	Yes, 16 years <first percentile	7/7
Dysmorphic features	High palate, oedematous eyelids and nasal bridge, wide nasal bridge, retrognathia	Buphthalmos, high broad forehead, mild lymphedema of hands and feet	Bitemporal narrowing, hirsutism, exotropia, exophthalmos, short nose, enlarged left eye with macrocornea	Oedematous eyelids, wide nasal bridge, retrognathia, epicanthal fold	Oedematous eyelids, dysplastic ears	Course facial features, downslanting palpebral	Course facial features, downslanting palpebral	Flattened midface, flattened nasal bridge with short rounded tip that is upturned with	7/7

	Ng et al ⁹	Ng et al ⁹	Ng et al ⁹	CDG—0095	CDG—0096	CDG—0099	CDG—0108	Total
Gastrointestinal abnormalities	G-tube placement	G-tube placement	Feeding difficulties with reflux	Feeding difficulties and constipation	Feeding difficulties	NG-tube placement	G-tube placement	7/7
Respiratory abnormalities	Tracheostomy	Reactive airway disease	Recurrent bronchopneumonia	Recurrent pneumonia	Recurrent pneumonia	Recurrent pulmonary infections	Chronic lung disease	7/7
IUGR	Yes (<second percentile)	Yes (third percentile)	Yes (third percentile)	No	No	Yes	No	4/7
Pregnancy complications	Polyhydramnios	Polyhydramnios	Normal	Polyhydramnios	Normal	Normal	Normal	3/7
Ocular abnormalities	No	Congenital glaucoma	No	No	No	Congenital glaucoma	Cortical visual impairment, high myopia, intermittent esotropia	2/7
Endocrine abnormalities	Hypothyroidism	Not reported	No	No	Hypothyroidism	No	No	2/7
Renal abnormalities	Nephrocalcinosis	No	No	No	Renal stones	No	No	2/7
Cardiac abnormalities	Arrial septal defect, repaired; patent ductus arteriosus ligation	No	No	No	No	No	No	1/7

Abbreviations: WES, whole exome sequencing; WGS, whole genome sequencing.

Article 2



International Journal of
Molecular Sciences



Article

A Comprehensive Bioinformatics Analysis of *UBE2C* in Cancers

Hassan Dastsooz ¹, Matteo Cereda ², Daniela Donna ¹ and Salvatore Oliviero ^{1,2,*}

¹ Dipartimento di Scienze della Vita e Biologia dei Sistemi, Università di Torino, Via Accademia Albertina, 13, 10123 Torino, Italy; hassan.dastsooz@unito.it (H.D.); daniela.donna@unito.it (D.D.)

² Italian Institute for Genomic Medicine (IIGM), Via Nizza 52, 10126 Torino, Italy; matteo.cereda@iigm.it

* Correspondence: salvatore.oliviero@unito.it

Received: 1 April 2019; Accepted: 3 May 2019; Published: 7 May 2019



Abstract: Ubiquitination is one of the main post-translational modification of proteins. It plays key roles in a broad range of cellular functions, including protein degradation, protein interactions, and subcellular location. In the ubiquitination system, different proteins are involved and their dysregulation can lead to various human diseases, including cancers. By using data available from the Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) databases, we here show that the ubiquitin conjugating enzyme, E2C (*UBE2C*), is overexpressed in all 27 cancers we investigated. *UBE2C* expression is significantly higher in late-stage tumors, which might indicate its involvement in tumor progression and invasion. This study also revealed that patients with higher *UBE2C* levels showed a shorter overall survival (OS) time and worse OS prognosis. Moreover, our data show that *UBE2C* higher-expression leads to worse disease-free survival prognosis (DFS), indicating that *UBE2C* overexpression correlates with poor clinical outcomes. We also identified genes with positive correlations with *UBE2C* in several cancers. We found a number of poorly studied genes (family with sequence similarity 72-member D, *FAM72D*; meiotic nuclear divisions 1, *MND1*; mitochondrial fission regulator 2, *MTFR2*; and POC1 centriolar protein A, *POC1A*) whose expression correlates with *UBE2C*. These genes might be considered as new targets for cancers therapies since they showed overexpression in several cancers and correlate with worse OS prognosis.

Keywords: *UBE2C*; TCGA; GEPIA; GTEx; UALCAN; overexpression; cancer

1. Introduction

Most proteins post-translational modifications are essential for proper cellular localization, substrate activity, and associations with other proteins. One evolutionarily conserved modification is ubiquitination, which can involve one or multiple defined lysine (K) residues or the N-terminal methionine residue (M1) [1–3].

Ubiquitin, a highly conserved regulatory protein (76 amino acids), plays an essential role in modulating protein functions. Ubiquitination is a post-translational modification mediated by a multi-step process which involves three different enzymes, including E1 ubiquitin-activating enzymes, E2 ubiquitin-conjugating enzymes, and E3 ubiquitin-ligating enzymes. This function leads to the proteasomal elimination of its substrate or alteration of the substrate activity, localization, and associations with other partners in its protein networks [2,4,5].

In the ubiquitin-proteasome pathway, anaphase promoting complex/cyclosome (APC/C) and the ubiquitin conjugating enzyme, E2C (*UBE2C*), are involved in the initiation of ubiquitin chain formation on APC/C substrates. *UBE2C* principally create Lys-11 (K11)-linked polyubiquitination on these substrates and then APC/C and another E2 enzyme, *UBE2S*, elongates and branches the ubiquitin, making more efficient proteolytic degradation signals (i.e., on mitotic cyclins) for the proteasome receptor, S5A, regulating

mitosis progression. Various cellular processes are regulated by the ubiquitin system; therefore, it is expected its dysregulation results in human diseases, including cancers [6,7]. *UBE2C* overexpression has been found in different human cancers, including hepatocellular carcinoma [8], thyroid [9], colon [10], breast [11], lung [12], brain [13], and cervical cancer [14].

It has been shown that *UBE2C* overexpression leads to chromosomes mis-segregation and alters the cell cycle process, facilitating cell proliferation [15,16]. Moreover, it has also been reported that *UBE2C* overexpression correlates with tumor progression and poor prognosis in many tumors [9,16–19]. In this study, the expression levels of *UBE2C* were evaluated in 27 different cancers using data from the Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) databases. We provide evidences that *UBE2C* acts as a proto-oncogene and can be considered as a therapeutic target for most cancers. Our results indicate that *UBE2C* is overexpressed in 27 studied cancers and its overexpression correlates worsen the overall survival (OS), suggesting its involvement in tumor progression and invasion. Our study also identified a number of genes that are in the *UBE2C* regulatory network.

2. Results

2.1. *UBE2C* Overexpression in Tumors, Their Pathological Stages, and Subtypes

Data extracted from TCGA database revealed that *UBE2C* expression was notably higher in all 27 tumor types compared to matched TCGA normal tissues and GTEx data (Figure 1). We next assessed the expression of *UBE2C* in normal tissue using RNA-sequencing data available from GTEx data. In particular, we compared expression levels of *UBE2C* between tumors with respect to normal matches, and data of GTEx. We found that *UBE2C* showed increased levels in all these cancers with respect to its expression in the normal tissues. The significant differences between all tumors and normal samples as a boxplot are given individually in Supplementary data Figure S1.

We next assessed the expression levels of *UBE2C* with respect to the molecular and histological subtypes of tumors, tumor grades, and other patient conditions when data are available using UALCAN.

In urologic cancers, we found that histological subtypes of BLCA show increased expression in both papillary and non-papillary tumors compared to normal (Table 1 and Figure S2 panel 1A). In relation to its molecular subtype, all reveal upregulated compared to normal with more statistically significant values for luminal papillary, followed by basal squamous (Table 1 and Figure S2 panel 1B). In renal cancers, KIRC shows increased expression levels of *UBE2C* in all grades compared to normal, with more significant for grade 2, followed by grade 3 and 4 (Table 1 and Figure S2 panel 1C). For its subtypes, both clear cell type A (ccA) and B (ccB) (p -value $< 10^{-8}$) subtypes show higher expression compared to normal, with slightly more significant for ccB (Table 1 and Figure S2 panel 1D). In KIRP tumors, all histological tumor subtypes showed *UBE2C* overexpression with high significance for type1 papillary renal cell carcinoma (RCC), followed by type2 papillary RCC (Table 1 and Figure S2 panel 1E). In PRAD tumors, the increase was statistically more significant for Gleason score 8, followed by Gleason score 9 and Gleason score 7 (Table 1 and Figure S2 panel 1F). The more statistically significant molecular signature was observed for erythroblast transformation-specific (ETS) transcription factor *ERG* (*ERG*) fusion, speckle type BTB/POZ protein (*SPOP*) mutation, and ETS translocation variant 1 (*ETV1*) fusion (Table 1 and Figure S2 panel 1G). In relation to the expression of *UBE2C* in metastatic PRAD based on androgen receptor (*AR*) amplification and *ERG* fusion, there is no significance difference compared to condition without these changes (Table 1 and Figure S2 panel 1H).

Compared to normal tissues in BRCA tumors, the expression of *UBE2C* was higher in all different subtypes, including triple negative breast cancer (TNBC), *HER2*-amplification, and luminal subtype (Table 1 and Figure S2 panel 2A). For its TNBC types, the statistically significant changes were seen in TNBC-mesenchymal (M), followed by TNBC-immunomodulatory (IM), TNBC-basal-like2 (BL2), and TNBC-basal-like1 (BL1; Table 1 and Figure S2 panel 2B). The *UBE2C* expression in this BRCA was increased in all pre-, peri-, and post-menopause conditions compared to the normal tissue, but not significant compared to each other (Table 1 and Figure S2 panel 2C). In addition, *UBE2C* expression in BRCA showed high levels in all histological subtypes, with the most significant increase in infiltrating lobular carcinoma (ILC) and infiltrating ductal carcinoma (IDC; Table 1 and Figure S2 panel 2D). The expression of amplified *MYC* proto-oncogene (*MYC*), cyclin D1 (*CCND1*), and Erb-B2 receptor tyrosine kinase 2 (*ERBB2*) in metastatic breast cancer compared to conditions without amplification indicated no significant correlation with *UBE2C* expression (Table 1 and Figure S2 panel 2E).

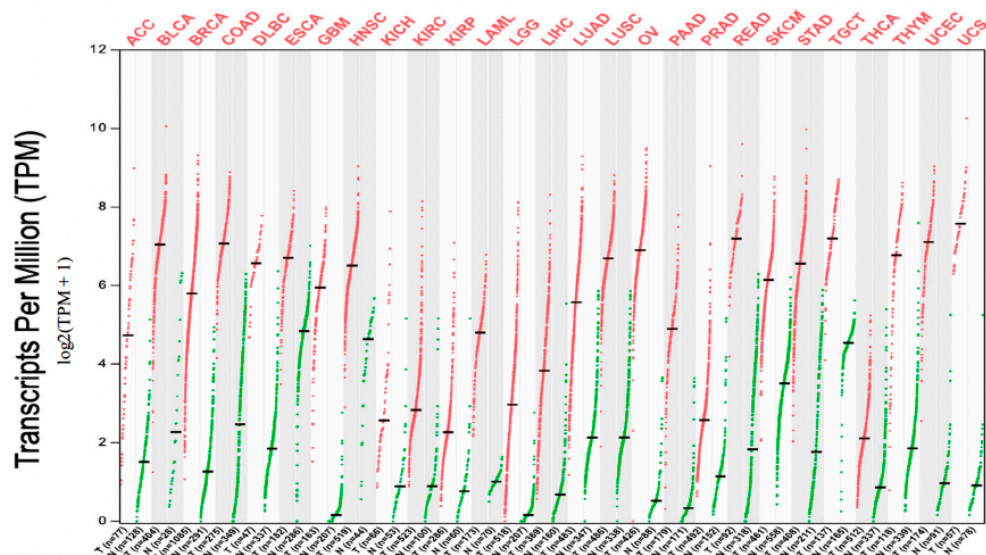


Figure 1. *UBE2C* expression in cancers. Expression level of *UBE2C* across 27 TCGA tumors compared to TCGA normal and GTEx data using GEPIA (Gene Expression Profiling Interactive Analysis) webserver. It is clear that in all 27 cancers there is notable upregulation of this gene. For each TCGA tumor (red), its matched normal and GTEx data (green) are given; T: tumor; N: normal; n: number. Y axis: transcript per million ($\log_2(\text{TPM} + 1)$). X axis: number of tumor and normal samples. ACC: adrenocortical carcinoma; BLCA: bladder urothelial carcinoma; BRCA: breast invasive carcinoma; COAD: colon adenocarcinoma; DLBC: lymphoid neoplasm diffuse large B-cell lymphoma; ESCA: esophageal carcinoma; GBM: glioblastoma multiforme; HNSC: head and neck squamous cell carcinoma; KICH: kidney chromophobe; KIRC: kidney renal clear cell carcinoma; KIRP: kidney renal papillary cell carcinoma; LAML: acute myeloid leukemia; LGG: brain lower grade glioma; LIHC: liver hepatocellular carcinoma; LUAD: lung adenocarcinoma; LUSC: lung squamous cell carcinoma; OV: ovarian serous cystadenocarcinoma; PAAD: pancreatic adenocarcinoma; PRAD: prostate adenocarcinoma; READ: rectum adenocarcinoma; SKCM: skin cutaneous melanoma; STAD: stomach adenocarcinoma; TGCT: testicular germ cell tumors; THCA: thyroid carcinoma; THYM: thymoma; UCEC: uterine corpus endometrial carcinoma; UCS: uterine carcinosarcoma.

Table 1. Statistically significant UBE2C overexpression based on histological, molecular subtypes, and different patient statuses (only findings with p -value < 0.05 are given).

Tumor	Histological Subtypes	Molecular Subtypes	Tumor Grade	Other Patient Conditions
BLCA	N-vs.-Papillary tumors: $p < 10^{-12}$ N-vs.-NonPapillary tumors: $p < 10^{-12}$	N-vs.-Neuronal: $p < 10^{-8}$; N-vs.-Basal squamous: $p < 10^{-12}$; N-vs.-Luminal: $p < 10^{-10}$; N-vs.-Luminal_Infiltrated: $p < 10^{-9}$; N-vs.-Luminal_Papillary: $p < 10^{-12}$		
KIRC		N-vs.-ccA subtype: $p = 10^{-8}$ N-vs.-ccB subtype: $p < 10^{-8}$	N-vs.-G 1: $p < 10^{-3}$ N-vs.-G 2: $p < 10^{-11}$ N-vs.-G 3: $p < 10^{-9}$ N-vs.-G 4: $p < 10^{-6}$	
KIRP	N-VS.-Type1 PRCC: $p < 10^{-12}$ N-VS.-Type2 PRCC: $p < 10^{-7}$ N-VS.-KIRP CIMF: $p < 10^{-3}$ N-VS.-Unclassified PRCC: $p < 10^{-3}$			
PRAD		N-vs.-ERG fusion: $p < 10^{-5}$ N-vs.-ETV1 fusion: $p < 10^{-3}$ N-vs.-SFOP mutation: $p < 10^{-5}$		N-vs.-Gleason score 7: $p < 10^{-3}$ N-vs.-Gleason score 8: $p < 10^{-5}$ N-vs.-Gleason score 9: $p < 10^{-4}$
BRCA	N-vs.-IDC: $p < 10^{-12}$ N-vs.-ILC: $p < 10^{-12}$ N-vs.-Mixed: $p < 10^{-6}$ N-vs.-Other: $p < 10^{-7}$ N-vs.-Mucinous: $p < 10^{-3}$ N-vs.-Metaplastic: $p < 10^{-3}$ N-vs.-Medullary: $p < 10^{-3}$	N-vs.-Luminal: $p < 10^{-12}$ N-vs.-HER2 Positive: $p < 10^{-10}$ N-vs.-TNBC: $p < 10^{-12}$ N-vs.-TNBC-BL1: $p < 10^{-7}$ N-vs.-TNBC-BL2: $p < 10^{-8}$ N-vs.-TNBC-IM: $p < 10^{-8}$ N-vs.-TNBC-LAR: $p < 10^{-3}$ N-vs.-TNBC-MSL: $p < 10^{-4}$ N-vs.-TNBC-M: $p < 10^{-10}$ N-vs.-TNBC-UNS: $p < 10^{-5}$		N-vs.-Pre-Menopause: $p < 10^{-12}$ N-vs.-Peri-Menopause: $p < 10^{-6}$ N-vs.-Post-Menopause: $p < 10^{-12}$
COAD	N-vs.-Adenocarcinoma: $p < 10^{-12}$ N-vs.-Mucinous-adenocarcinoma: $p < 10^{-11}$			

Table 1. Cont.

Tumor	Histological Subtypes	Molecular Subtypes	Tumor Grade	Other Patient Conditions
ESCA	N-vs.-Adenocarcinoma: $p < 10^{-12}$ N-vs.-Squamous-cell-carcinoma: $p < 10^{-12}$		G 2-vs.-G 3: $p < 10^{-8}$	
HNSC			N-vs.-G 1: $p < 10^{-12}$ N-vs.-G 2: $p < 10^{-12}$ N-vs.-G 3: $p < 10^{-12}$ N-vs.-G 4: $p < 10^{-5}$	N-vs.-HPV+ve: $p < 10^{-11}$ N-vs.-HPV-ve: $p < 10^{-12}$
LIHC			N-vs.-G 1: $p < 10^{-5}$ N-vs.-G 2: $p < 10^{-12}$ N-vs.-G 3: $p < 10^{-12}$ N-vs.-G 4: $p < 10^{-3}$	
PAAD			G1-vs.-G 2: $p = 0.014$ G1-vs.-G3: $p = 0.0062$	N-vs.-Weekly Drinker: $p = 0.025$ N-vs.-Occasional Drinker: $p < 10^{-3}$ N-vs.-Diabetic: $p < 10^{-4}$ N-vs.-NonDiabetic: $p < 10^{-7}$ N-vs.-Pancreatitis: $p < 10^{-4}$ N-vs.-NoPancreatitis: $p < 10^{-7}$
READ	N-vs.-Adenocarcinoma: $p < 10^{-12}$ N-vs.-Mucinous-adenocarcinoma: $p < 10^{-3}$			
STAD	N-vs.-Adenocarcinoma (NOS): $p < 10^{-12}$ N-vs.-Adenocarcinoma (Diffuse): $p < 10^{-10}$ N-vs.-Adenocarcinoma (Signet Ring): $p < 10^{-3}$ N-vs.-Intestinal Adenocarcinoma (NOS): $p < 10^{-12}$ N-vs.-IntestinalAdenocarcinoma (Tubular): $p < 10^{-12}$ N-vs.-IntestinalAdenocarcinoma (Mucinous): $p < 10^{-4}$ N-vs.-IntestinalAdenocarcinoma (Papillary): $p = 0.017$		N-vs.-G 1: $p = 0.0057$ N-vs.-G 2: $p < 10^{-12}$ N-vs.-G 3: $p < 10^{-12}$ G 1-vs.-G 3: $p = 0$	N-vs.-Tumors (with <i>H. pylori</i> infection): $p < 10^{-4}$ N-vs.-Tumors (without <i>H. pylori</i> infection): $p < 10^{-12}$ N-vs.-Tumors (Not available): $p < 10^{-12}$

Table 1. Cont.

Tumor	Histological Subtypes	Molecular Subtypes	Tumor Grade	Other Patient Conditions
LUAD	N-vs.-NOS: $p < 10^{-12}$			
	N-vs.-Mixed: $p < 10^{-14}$			
	N-vs.-LBC-NonMucinous: $p < 10^{-4}$			
	N-vs.-SolidPatternPredominant: $p = 0.019$			
	N-vs.-Acinar: $p < 10^{-5}$			
	N-vs.-LBC-Mucinous: $p < 10^{-5}$			
LUSC	N-vs.-Mucinous carcinoma: $p < 10^{-3}$			
	N-vs.-Papillary: $p < 10^{-4}$			
	N-vs.-NOS: $p < 10^{-12}$			
UCEC	N-vs.-Basaloid: $p < 10^{-5}$			
	N-vs.-Papillary: $p = 0.018$			
	N-vs.-Endometrioid: $p < 10^{-12}$			N-vs.-Pre-Menopause: $p < 10^{-9}$
THCA	N-vs.-Serous: $p < 10^{-12}$			N-vs.-Peri-Menopause: $p < 10^{-4}$
	N-vs.-Mixed serous and endometrioid: $p < 10^{-7}$			N-vs.-Post-Menopause: $p < 10^{-12}$
	Classical-VS.-Follicular: $p < 10^{-5}$			
	Tall-VS.-Follicular: $p < 10^{-3}$			

N: normal; p : p -value; amp: amplification; G: grade. All details are given in Supplementary Figure S2.

In relation to digestive system tumors, COAD tumors showed increased *UBE2C* levels in adenocarcinoma and mucinous-adenocarcinoma (Table 1 and Figure S2 panel 3A). In ESCA, *UBE2C* expression was increased in both subtypes, including adenocarcinoma and squamous-cell-carcinoma (Table 1 and Figure S2 panel 3B). The expression in HNSC in all grades is higher than in normal tissue, particularly for grade 3 tumors (Table 1 and Figure S2 panel 3C). Also, the expression based on human papilloma virus (HPV) status showed a more statistically significant increase in HPV negatives than positives compared to normal in this cancer (Table 1 and Figure S2 panel 3D). All grades of LIHC tumors showed high *UBE2C*, with grade 2 and 3 more statically significant (Table 1 and Figure S2 panel 3E). In PAAD tumors based on patients' drinking habits, the expression was more statistically significant only for occasional drinkers and weekly drinkers (Table 1 and Figure S2 panel 3F). On the basis of diabetes status, the expression in non-diabetics was more significant than diabetics when compared to normal (Table 1 and Figure S2 panel 3G), but their comparison with each other was not statistically significant. On the basis of pancreatitis status, the expression was more significant in non-pancreatitis than pancreatitis compared to their matched normal (Table 1 and Figure S2 panel 3H), but the expression was not significant when these two were compared to each other. In READ, both adenocarcinoma and mucinous-adenocarcinoma showed a significant increase of expression, with the first one more statistically significant (Table 1 and Figure S2 panel 3I). STAD tumors showed high *UBE2C* expression in all conditions with high significance in tumors without *H. pylori* infection (Table 1 and Figure S2 panel 3J), but the comparison with each other was not statistically significant. For its histological subtypes, all of them showed increased levels, with more statistically significant for intestinal adenocarcinoma-not otherwise specified (NOS), intestinal adenocarcinoma tubular, and adenocarcinoma NOS with the same change, followed by adenocarcinoma diffuse and intestinal adenocarcinoma mucinous (Table 1 and Figure S2 panel 3K). Also, all grades of STAD show significant *UBE2C* overexpression compared to normal (Table 1).

Regarding lung cancers, while approximately all histological subtypes of LUAD tumors showed increased *UBE2C* expression, the increased level was more statistically significant for lung adenocarcinoma mixed type, followed by lung adenocarcinoma NOS (Table 1 and Figure S2 panel 4A). Regarding LUSC, for its histological subtypes, the increase was more significant for LUSC NOS, followed by lung basaloid squamous cell carcinoma (Table 1 and Figure S2 panel 4B).

All histological subtypes of UCEC tumors showed a notable *UBE2C* increase, which was more statistically significant for serous and endometrioid and then mixed serous and endometrioid (Table 1 and Figure S2 panel 5A). For this cancer type, the *UBE2C* expression was more significant for post-menopause (Table 1 and Figure S2 panel 5B).

When we inspected the contribution of smoking habits of cancer patients to *UBE2C* expression, we found no difference between smokers and non-smokers in BLCA patients. Nevertheless, the *UBE2C* expression level was higher in reformed smokers (>15 years) compared to non-smokers (Figure S2 panel 6A). Regarding smoking habits in ESCA, the expression levels in all conditions were higher than normal, but was more significant in smokers than non-smokers (Figure S2 panel 6B). In LUAD, the expression in smokers (reformed smoker2, smokers, and reformed smoker1) showed more significant values than non-smokers (Figure S2 panel 6C). Regarding LUSC patients based on smoking habits, the expression in all categories showed an increase compared to normal, which was more significant for reformed smoker1, smoker, and reformed smoker2. However, there was no statistically significant difference between smokers and non-smokers in LUSC cancer (Figure S2 panel 6D).

We next investigated *UBE2C* expression on the basis of patients' pathological stage in TCGA cancer types. We found that in COAD, ESCA, HNSC, KICH, READ, STAD, and BLCA, *UBE2C* expression levels were significantly higher in early-stages (Figure 2, p -value < 0.05). This indicates a possible involvement of *UBE2C* in the initiation of cancer. Furthermore, the expression in BRCA, KIRC, KIRP, LIHC, LUAD, LUSC, and UCEC was higher in late-stage cancers compared to early stages, representing a possible role of *UBE2C* in cancer progression and invasion (Figure 2, cancer without and/or small numbers of normal matches (when there is only one sample in each stage) were excluded from this analysis).

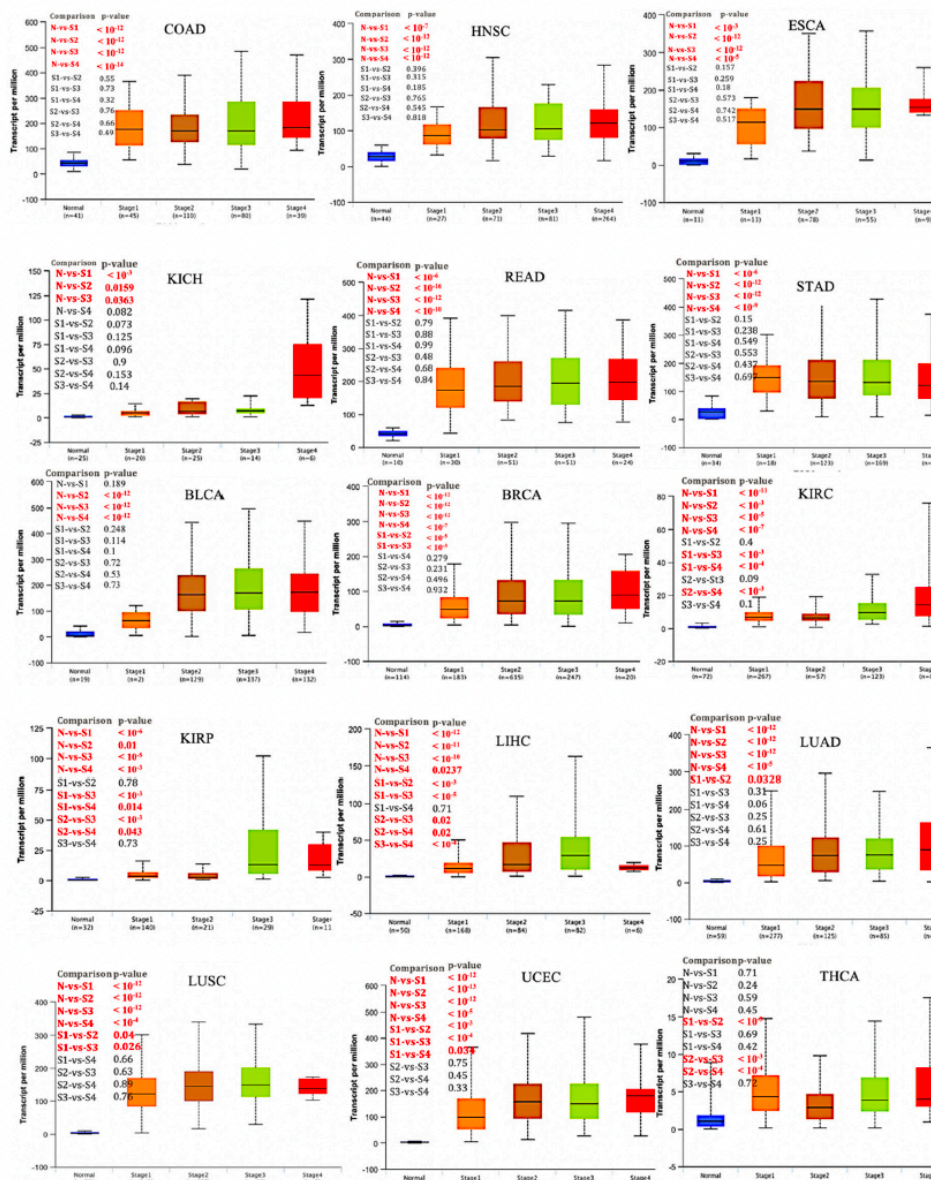


Figure 2. *UBE2C* expression based on individual pathological cancer stage. Box plot reveals that the over-expression of *UBE2C* may have role in initiation of COAD, HNSC, ESCA, KICH, READ, STAD, and BLCA, but not in progression since significant changes were observed only between normal and pathological stages not between each stage. The expression of *UBE2C* in BRCA, KIRC, KIRP, LIHC, LUAD, LUSC, and UCEC shows its involvement in both cancer initiation and progression. Regarding THCA, while there were no significant changes between normal and pathological stages, it seems that *UBE2C* is involved in progression from stage 1 to 2 and then 2 to 3. Y axis: transcript per million, X axis: pathological cancer stages with the number of samples in each stage in parenthesis. N: normal, S: stage.

2.2. Role of *UBE2C* Overexpression in Cancer Prognosis

The OS time between *UBE2C* higher-expression-level and *UBE2C* lower-expression-level tumors were compared in TCGA tumor types and data revealed a shorter OS with worse prognosis in patients with *UBE2C* higher expression levels compared to its lower expression levels in the following cancers:

ACC, BRCA, KIRC, KIRP, LGG, LUAD, PAAD, and SKCM (Figure 3, only cancers with significant changes, p -value < 0.05, a shorter OS with worse prognosis, are given).

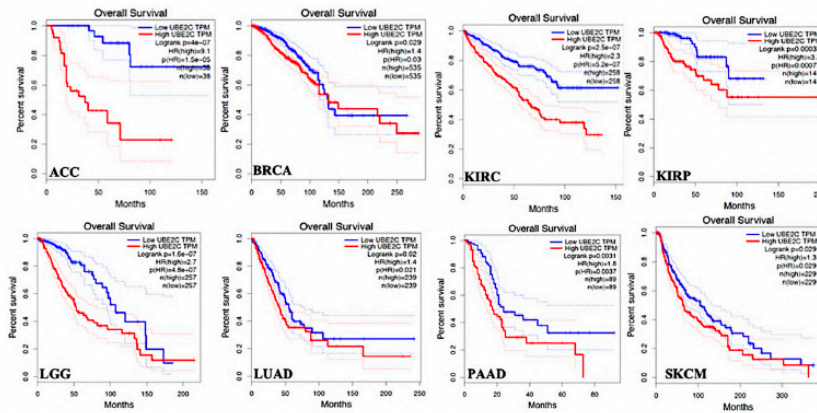


Figure 3. OS time between *UBE2C* higher-expression-level and *UBE2C* lower-expression-level tumors in TCGA tumor types with shorter overall survival time and worse OS prognosis. Red line shows the cases with highly expressed *UBE2C* and blue line is indicated for the cases with lowly expressed *UBE2C*. HR: hazard ratio.

Regarding DFS time in the TCGA tumor types, data showed that *UBE2C* higher-expression levels led to worse DFS prognosis in comparison to its lower expression in the following tumors: ACC, KIRC, KIRP, LGG, LIHC, PAAD, PRAD, THCA, and UCEC. (Figure 4, only cancers with significant changes, p -value < 0.05, worse DFS prognosis, are given). These data demonstrate that *UBE2C* overexpression results in poor clinical outcomes in the above tumors.

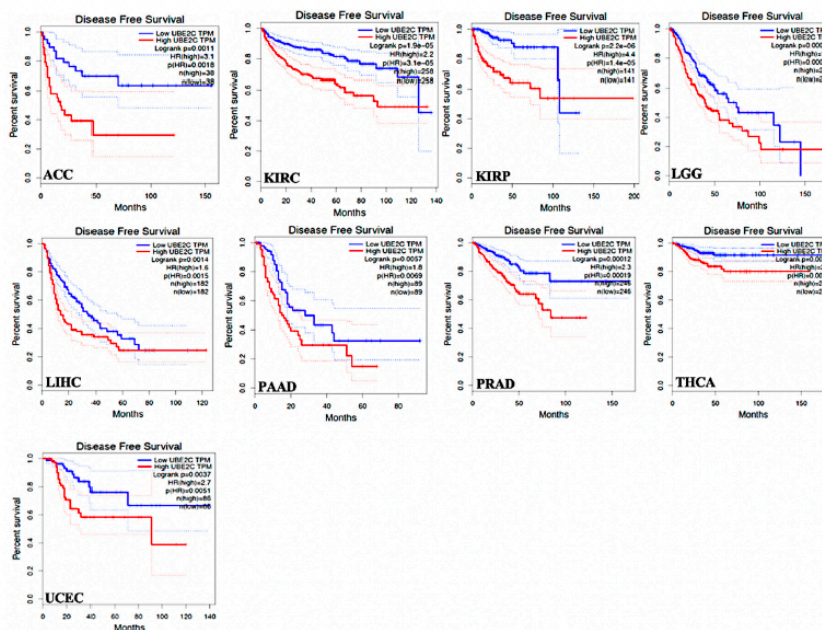


Figure 4. DFS time between *UBE2C* higher-expression-level and *UBE2C* lower-expression-level tumors in the TCGA tumor types with worse prognosis. Red line shows the cases with highly expressed *UBE2C* and blue line is indicated for the cases with lowly expressed *UBE2C*. HR: hazard ratio.

2.3. Gene Expression Correlation between UBE2C and Other Genes in Cancers

Our study revealed that the expression of *UBE2C* has a moderate to very strong positive correlation with other genes in 27 cancers (Supplementary Table S1 file, all information related to coefficient correlation and *p*-value are given in this file and different colors are used to distinguish the correlation as follows: Strong and very strong positive correlation in green; medium positive correlation in black, weak and very weak positive correlation in red; and negative correlation in violet). As seen in Supplementary Table S1 file and Table 2, the positive *UBE2C* expression correlations are strong to very strong (R between 0.6 and 1 and *p*-value < 0.05) for the following genes in all 27 cancers: MYB proto-oncogene like 2 (*MYBL2*), trophinin associated protein (*TROAP*), cell division cycle 20 (*CDC20*), centromere protein A (*CENPA*), kinesin family member C1 (*KIFC1*), cyclin dependent kinase 1 (*CDK1*), kinesin family member 4A (*KIF4A*), and kinesin family member 20A (*KIF20A*). In addition, the following genes showed a strong to very strong positive expression in correlations with *UBE2C* in 26 cancers, but moderate positive correlations (R between 0.4 and 0.59 and *p*-value < 0.05, Table 2 and Table S1 file) in one cancer, including TPX2, microtubule nucleation factor (*TPX2*), polo like kinase 1 (*PLK1*), aurora kinase B (*AURKB*), non-SMC condensin I complex subunit G (*NCAPG*), cyclin B1 (*CCNB1*), spindle and kinetochore associated complex subunit 3 (*SKA3*), and kinesin family member 18B (*KIF18B*). Moreover, some genes showed strong to very strong positive expression correlations with *UBE2C* in several cancers while moderate correlations were found in a few cancers among 27 studied cancers as indicated in the Table S1 file and Table 2. Among the 27 cancers, most negative correlations between *UBE2C* expression and different genes were observed in the TGCT cancer mentioned in Table S1 file (in violet) and Table 2, with a strong negative correlation for Testis-specific Y-encoded-like protein 2 (*TSPYL2*), ATR serine/threonine kinase (*ATR*), and CYLD lysine 63 deubiquitinase (*CYLD*).

Regarding genes which have protein products that have transcription factor binding sites on both the promoter and enhancer regions of *UBE2C*, we found forkhead box M1 (*FOXM1*, two sites with transcription start site (TSS) distance of +0.1 kb and +158.3 kb), E2F transcription factor 1 (*E2F1*, TSS distance: +158.3 kb), RAD51 recombinase (*RAD51*, two sites with TSS distance of +551.6 kb and +123.8 kb), and BRCA1 DNA repair associated (*BRCA1*, TSS distance: +158.3 kb). It is worthwhile to note that only *FOXM1*, *E2F1*, and *RAD51* showed positive correlations with *UBE2C* in all 27 cancers. These three genes showed similar strong to very strong positive correlations with *UBE2C* in the different cancers given in Table 2 and Supplementary data Table S1.

In relation to the expression correlation between *UBE2C* and tumor suppresser genes including BUB1 mitotic checkpoint serine/threonine kinase B (*BUB1B*), *BRCA1*, BRCA2 DNA repair associated (*BRCA2*), checkpoint kinase 2 (*CHK2*), ATM serine/threonine kinase (*ATM*), *ATR*, tumor protein p53 (*TP53*), *CYLD*, and *TSPYL2*, the positive correlation between *BUB1B* and *UBE2C* was seen in 26 cancers, but not in TGCT cancer (Table 2 and Supplementary Data Table S1). Regarding *BRCA1* and *BRCA2*, both of these genes showed similar strong positive correlations with *UBE2C* in different cancers (Table 2 and Table S1 file), but a negative correlation was observed for *BRCA1* in THCA and for *BRCA2* in TGCT (Table 2 and Supplementary Data Table S1). *CHK2* and *TP53* genes showed a positive correlation with *UBE2C* in several cancers (Table 2 and Table S1 file) while *CHK2* showed a strong negative correlation in KICH and *TP53* revealed a very weak negative correlation in this cancer. Three tumor suppresser genes, including *CYLD*, *ATM*, and *TSPYL2*, showed a negative correlation with *UBE2C* in several cancers with a similar moderate to strong negative correlation for them in the following cancers: LUSC, READ, UCEC, OV, and UCS. It is worthy to note that *TSPYL2* showed down-regulation in most of the 27 cancers (Supplementary Data Figure S3) and also very weak to strong negative correlation with *UBE2C* in most of these studied cancers (Table S1 file and Table 2). Among the genes with a positive expression correlation with *UBE2C*, some genes that can be considered as target genes in several cancers. For example, mitochondrial fission regulator 2 (*MTFR2* also called *FAM54A*, only limited studies in GBM [20] and ovarian cancer [21]), meiotic nuclear divisions 1 (*MND1*, limited studies in breast [22] and ovarian cancers [23]), family with sequence similarity 72 member D (*FAM72D* only one study in GBM [24]), and POC1 centriolar protein A (*POC1A*, limited reports in bladder [25], brain [26], and breast cancer [27]). As seen in Table 3 and Supplementary Figure S4 file, they showed an overexpression, with alterations across different pathological cancer stages and a worse OS prognosis in several cancers.

Table 2. Expression correlation between UBE2C and various genes.

Genes with Strong to very Strong Positive UBE2C Expression Correlations in all 27 Cancers	Genes with Strong to very Strong Positive UBE2C Expression Correlations in 26 Cancers, but Moderate Positive Correlation in One Cancer	Genes with Strong to very Strong Positive UBE2C Expression Correlations in Several Cancers, but Moderate in Few Cancers	Gene with Negative Correlation with UBE2C Expression in TGCT Cancer
MYBL2, TROAP, CDC20, CENPA, KIFCL, CDK1, KIF4A, and KIF20A.	TPX2, PLK1, AURKB, NCAPG, CCNB1, SKA3, and KIF18B.	CDCA3, HJURP, UBE2T, MAD2L1, DLGAP5, MELK, RNASEH2A, KIF23, FAM72D, and FEN1	KIF2C, CDCA5, CDC25C, BUB1, MTFR2, KIF15, CDKN3, CKS2, BUB1B, ASF1B, NUF2, PLK4, TTK, POC1A, UBE2S, AUNIP, NEK2, DSN1, BRCA2, CYLD, ATM, ATR, and TSPYL2
Gene with negative correlation with UBE2C expression in different cancers			Positive expression correlation between UBE2C and genes having TF binding sites on both promoter and enhancer regions of UBE2C in different cancers
<p>BRCA1 in THCA</p> <p>BRCA2 in TGCT</p> <p>CHK2 in KICH, and THYM</p> <p>CYLD in COAD, KIRP, LUAD, LUSC, PRAD, READ, SKCM, STAD, THCA, UCEC, ACC, LAML, OV, UCS, THYM, GBM, TGCT</p> <p>ATM in BLCA, BRCA, COAD, KIRC, LUAD, LUSC, PAAD, READ, STAD, THCA, UCEC, ACC, OV, UCS, THYM, GBM, HNSC, TGCT</p> <p>ATR in LUAD, SKCM, THCA, UCEC, ACC, OV, UCS, TGCT</p> <p>TP53 in KICH</p> <p>TSPYL2 in BLCA, BRCA, COAD, ESCA, KICH, LIHC, LUAD, LUSC, PAAD, PRAD, READ, SKCM, STAD, THCA, UCEC, ACC, OV, UCS, THYM, LGG, GBM, HNSC, TGCT</p>			<p>BRCA (FOXMI: R = 0.69; E2F1: R = 0.7; RAD51: R = 0.67)</p> <p>ESCA (FOXMI: R = 0.79; E2F1: R = 0.82; RAD51: R = 0.83)</p> <p>KICH (FOXMI: R = 0.84; E2F1: R = 0.83; RAD51: R = 0.63)</p> <p>KIRC (FOXMI: R = 0.73; E2F1: R = 0.7; RAD51: R = 0.72)</p> <p>LUSC (FOXMI: R = 0.72; E2F1: R = 0.67; RAD51: R = 0.7)</p> <p>PAAD (FOXMI: R = 0.82; E2F1: R = 0.88; RAD51: R = 0.86)</p> <p>READ (FOXMI: R = 0.73; E2F1: R = 0.86; RAD51: R = 0.74)</p> <p>SKCM (FOXMI: R = 0.76; E2F1: R = 0.7; RAD51: R = 0.67)</p> <p>UCEC (FOXMI: R = 0.7; E2F1: R = 0.75; RAD51: R = 0.66)</p> <p>ACC (FOXMI: R = 0.84; E2F1: R = 0.69; RAD51: R = 0.86)</p> <p>DLBC (FOXMI: R = 0.89; E2F1: R = 0.82; RAD51: R = 0.9)</p> <p>THYM (FOXMI: R = 0.72; E2F1: R = 0.84; RAD51: R = 0.82)</p> <p>LGG (FOXMI: R = 0.9; E2F1: R = 0.8; RAD51: R = 0.94)</p> <p>GBM (FOXMI: R = 0.87; E2F1: R = 0.75; RAD51: R = 0.87)</p>
Positive correlation between UBE2C and tumor suppressor genes in cancers			
<p>BUB1B: Strong and very strong (BLCA, BRCA, COAD, ESCA, KICH, KIRC, KIRP, LIHC, LUAD, LUSC, PAAD, READ, SKCM, STAD, UCEC, ACC, DLBC, LAML, UCS, THYM, GBM, HNSC); Moderate (PRAD, THCA, OV); Weak (LGG)</p> <p>BRCA1: Strong and very strong (COAD, ESCA, LUAD, LUSC, PAAD, READ, ACC, DLBC, LAML, LGG, GBM); Moderate (BLCA, KICH, KIRC, KIRP, LIHC, PRAD, SKCM, STAD, OV, UCS, THYM, HNSC); Weak (BRCA and UCEC); Very weak (TGCT)</p> <p>BRCA2: Strong and very strong (COAD, ESCA, READ, DLBC, LAML, LGG, GBM); Moderate (BLCA, BRCA, KIRP, LIHC, LUAD, LUSC, PAAD, SKCM, STAD, UCEC, ACC, OV, and UCS); Weak (KICH, KIRC, PRAD, THCA, HNSC); Very weak (THYM)</p> <p>CHK2: Strong and very strong (COAD, LUSC, PAAD, READ, UCEC, DLBC, LAML, LGG, GBM, TGCT); Moderate (BLCA, BRCA, ESCA, KIRC, LUAD, SKCM, STAD, UCS, HNSC); Weak (LIHC, PRAD, THCA, ACC); Very weak (KIRP and OV)</p> <p>TP53: Strong and very strong (ESCA, PAAD, READ, DLBC, LAML, THYM, LGG, GBM, TGCT); Moderate (COAD, LIHC, LUSC, UCS); Weak (KIRP, LUAD, STAD, THCA, UCEC, ACC); Very weak (BLCA, BRCA, KIRC, PRAD, SKCM, OV, HNSC)</p> <p>ATM: Weak (DLBC); Very weak (ESCA, KICH, KIRP, LIHC, PRAD, SKCM, DLBC, LGG)</p> <p>ATR: Strong and very strong (DLBC and LAML); Moderate (ESCA, PAAD, THYM); Weak (BLCA, COAD, KICH, KIRC, STAD, HNSC); Very weak (BRCA, KIRP, LIHC, LUSC, PRAD, READ, LGG, GBM)</p> <p>CYLD: Weak (PAAD, DLBC); Very weak (BLCA, BRCA, ESCA, KICH, LIHC, LGG, HNSC)</p>			

More details are given in Supplementary Table S1.

Table 3. Statistically significant overexpression of different proposed target genes in most common cancers.

Genes	Cancers with Overexpression of Proposed Cancer Target Genes (Normal-vs.-Primary)	OS (Higher Expression Levels-vs.-Lower Expression Levels)
MTFR2 (FAM54A)	BLCA (p-value < 10 ⁻¹⁶), BRCA (p-value < 10 ⁻¹²), Metastatic breast cancer (MYC-amplification (+) vs. MYC-amplification (-); p-value < 10 ⁻¹² , CCND1-amplification (+) vs. CCND1-amplification (-); p-value < 10 ⁻¹² , ERBB2-amplification (+) vs. ERBB2-amplification (-); p-value < 10 ⁻¹²), CESC (p-value < 10 ⁻¹²), COAD (p-value < 10 ⁻¹²), ESCA (p-value < 10 ⁻¹²), GBM (p-value < 10 ⁻¹²), HNSC (p-value < 10 ⁻¹²), KICH (p-value = 0.017), KIRC (p-value < 10 ⁻¹²), KIRP (p-value < 10 ⁻¹²), LIHC (p-value < 10 ⁻¹²), LUAD (p-value < 10 ⁻¹²), LUSC (p-value < 10 ⁻¹²), PRAD (p-value < 10 ⁻⁵ , Molecular subtypes of PRAD: Normal-vs.-ERG fusion: p-value < 10 ⁻³ , Normal-vs.-ETV1 fusion: p-value < 10 ⁻³ , Normal-vs.-SPOP mutation: p-value < 10 ⁻³), Metastatic prostate cancer (ERG-fusion (+) vs. ERG-fusion (-); p-value < 10 ⁻¹² , AR-amplification (+) vs. AR-amplification (-); p-value < 10 ⁻¹²), READ (p-value < 10 ⁻⁶), STAD (p-value < 10 ⁻¹²), UCEC (p-value < 10 ⁻¹²).	ACC (p-value < 0.0001) LGG (p-value < 0.0001) KICH (p-value < 0.0001) KIRC (p-value < 0.0001) KIRP (p-value < 0.0001) LIHC (p-value = 0.00018) UCEC (p-value = 0.043)
MND1	BLCA (p-value < 10 ⁻⁶), BRCA (p-value < 10 ⁻¹²), COAD (p-value < 10 ⁻¹²), ESCA (p-value < 10 ⁻¹²), CESC (p-value = p-value < 10 ⁻¹²), HNSC (p-value < 10 ⁻¹²), KICH (p-value < 10 ⁻⁹), KIRC (p-value < 10 ⁻¹²), KIRP (Normal-vs.-Primary: p-value < 10 ⁻¹⁵), LIHC (p-value < 10 ⁻¹²), LUAD (p-value < 10 ⁻¹²), LUSC (p-value < 10 ⁻¹²), PRAD (p-value < 10 ⁻⁸ , Molecular subtypes of PRAD: Normal-vs.-ERG fusion: p-value < 10 ⁻⁴ , Normal-vs.-ETV1 fusion: p-value < 10 ⁻³ , Normal-vs.-ETV4 fusion: p-value < 10 ⁻⁴), READ (p-value < 10 ⁻⁸), STAD (p-value < 10 ⁻¹²), THCA (p-value = 0.0348), UCEC (p-value < 10 ⁻¹²).	ACC (p-value = 0.016) LGG (p-value < 0.0001) KICH (p-value = 0.00025) KIRC (p-value = 0.046) KIRP (p-value < 0.0001) LIHC (p-value = 0.0054) LUAD (p-value = 0.0034) PAAD (p-value = 0.048) SKCM (p-value = 0.047)
FAM72D	BLCA (p-value < 10 ⁻¹²), BRCA (p-value < 10 ⁻¹²), Metastatic breast cancer (amplification (+) vs. ERBB2-amplification (-); p-value < 10 ⁻³), COAD (p-value < 10 ⁻¹²), ESCA (p-value < 10 ⁻¹²), GBM (p-value < 10 ⁻¹²), HNSC (p-value < 10 ⁻¹²), KIRC (p-value < 10 ⁻¹²), KIRP (p-value < 10 ⁻¹⁵), LIHC (p-value < 10 ⁻¹²), LUAD (p-value < 10 ⁻¹²), LUSC (p-value < 10 ⁻¹²), PRAD (p-value < 10 ⁻¹⁰ , Molecular subtypes of PRAD: Normal-vs.-ERG fusion: p-value < 10 ⁻⁷ , Normal-vs.-ETV1 fusion: p-value < 10 ⁻³ , Normal-vs.-SPOP mutation: p-value < 10 ⁻⁵), READ (p-value < 10 ⁻¹³), STAD (p-value < 10 ⁻¹²), UCEC (p-value < 10 ⁻¹²).	ACC (p-value < 0.0001) LGG (p-value < 0.0001) BRCA (p-value = 0.041) KICH (p-value = 0.0012) KIRC (p-value < 0.0001) KIRP (p-value < 0.0001) LIHC (p-value = 0.0024) LUAD (p-value = 0.009)
POCIA	BLCA (p-value < 10 ⁻¹⁰), BRCA (p-value < 10 ⁻¹²), Metastatic breast cancer (amplification (+) vs. ERBB2-amplification (-); p-value = 0.0279), CESC (p-value < 10 ⁻⁹), COAD (p-value < 10 ⁻¹²), ESCA (p-value < 10 ⁻¹¹), GBM (p-value < 10 ⁻¹²), HNSC (p-value < 10 ⁻¹²), KICH (p-value = 0.016), KIRP (p-value < 10 ⁻⁵), LIHC (p-value < 10 ⁻¹²), LUAD (p-value < 10 ⁻¹²), LUSC (p-value < 10 ⁻¹²), PRAD (p-value < 10 ⁻¹⁵ , Molecular subtypes of PRAD: Normal-vs.-ERG fusion: p-value < 10 ⁻⁷ , Normal-vs.-ETV1 fusion: p-value < 10 ⁻³ , Normal-vs.-ETV4 fusion: p-value < 10 ⁻² , Normal-vs.-SPOP mutation: p-value < 10 ⁻¹⁰), READ (p-value < 10 ⁻⁵), STAD (p-value < 10 ⁻¹⁶), THCA (p-value < 10 ⁻⁸), UCEC (p-value < 10 ⁻¹²).	ACC (p-value < 0.0001) LGG (p-value < 0.0001) KICH (p-value = 0.00017) KIRC (p-value < 0.0001) KIRP (p-value < 0.0001) LIHC (p-value = 0.028) LUAD (p-value = 0.0094) OV (p-value = 0.039) PAAD (p-value = 0.012) SKCM (p-value = 0.0039) STAD (p-value = 0.013)

2.4. UBE2C Protein Network

Genes with strong or very strong positive correlations with *UBE2C* expression in the 27 cancers (see Methods Section, Table 2, and Supplementary Table S1) and also some negative correlations with important tumor suppressor genes were identified in the *UBE2C* protein network. Data from STRING database revealed that all these proteins (products of all genes listed in Table S1 file) are in the same protein network (Figure 5, only proteins with strong and very strong positive correlations with *UBE2C* in most of the 27 cancers, some tumor suppressor proteins, and protein with TF bindings site on *UBE2C* are shown). Proteins in this network are involved in different pathways, including cell cycle, oocyte meiosis, p53 signaling pathway, double-strand break repair, oocyte development and differentiation, FoxO signaling pathway, ubiquitin mediated proteolysis, cellular senescence, and progesterone-mediated oocyte maturation among others (Table 4 and Supplementary data Table S2). All pathways that *UBE2C* is mainly involved with are indicated in red (Supplementary Table S2). Proteins with TF binding sites on both the promoter and enhancer regions of *UBE2C* (FOXM1, E2F1, RAD51, and BRCA1) are also involved in the *UBE2C* network. As described in Table 4 and in Supplementary data Table S2, these proteins were found to also be involved in most pathways related to the *UBE2C* protein network. Among the *UBE2C* protein partners, *TSPYL2* (a member of the testis-specific protein Y-encoded) is a tumor suppressor protein which acts in the chromatin remodeling process. This protein is also involved in most pathways with the involvement of *UBE2C* (Supplementary data Table S2). As *TSPYL2* was under-expressed in most cancers (Supplementary data Figure S3) and showed negative correlations with *UBE2C* that were very weak to moderate (Table 2 and Supplementary data Table S1), we analyzed the RNA–RNA association between *TSPYL2* and *UBE2C* using RNAup webserver and also RNA–protein interactions using RPISeq. These analyses showed that *TSPYL2* is not only involved in the *UBE2C* protein network, but also in RNA–RNA interactions and RNA–protein interactions with *UBE2C* (Figure S5 panel A and B). Moreover, these two proteins were in the same subcellular localization. We also found a D-box (one of the recognition amino acid sequences identified by APC/C in the ubiquitin–proteasome pathway) in amino acids 45 to 48 in *TSPYL2* protein using the GPS-ARM tool (Figure S5 panel C).

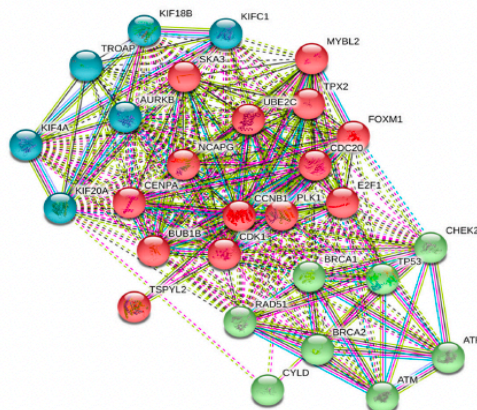


Figure 5. *UBE2C* protein network. Proteins with a strong to very strong positive correlation with *UBE2C* in most of the 27 cancers and some tumor suppressors are clustered in three categories based on the kmeans clustering option in STRING. In the cluster indicated in blue, different members of kinesin Family (KIF20A, KIF18B, KIFC1, and KIF4A) are included, which function as microtubule-dependent molecular motor, AURKB (a serine/threonine-protein kinase component of the chromosomal passenger complex, CPC) with an essential role in the regulation of mitosis, and TROAP involved in cell adhesion molecule complex. Another cluster shown in red (TPX2, *UBE2C*, PLK1, CDK1, CENPA, CDC20, MYBL2, BUB1B, CCNB1, NCAPG, SKA3, E2F1, FOXM1, and *TSPYL2*) is involved in different vital cellular processes, such as cell cycle, regulation of cellular metabolic process, cellular protein modification, signaling, chromosome organization, histone modification, and others as described in Table S2 file. The last cluster in green includes major tumor suppressor genes, including BRCA2, BRCA1, CHEK2, ATR, ATM, TP53, RAD51, and CYLD.

Table 4. Biological pathways related to proteins involved in UBE2C network.

KEGG Number	Cellular Process	Protein Gene
hsa04110	Cell cycle	ATM, ATR, BUB1, BUB1B, CCNA2, CCNB1, CCNB2, CDC20, CDC25A, CDC25C, CDK1, CHEK1, E2F1, MAD2L1, ORC6, PLK1, TP53, TTK, CHEK2
hsa04114	Oocyte meiosis	AURKA, BUB1, CCNB1, CCNB2, CDC20, CDC25C, CDK1, MAD2L1, PLK1, SGOL1
hsa04914	Progesterone-mediated oocyte maturation	AURKA, BUB1, CCNA2, CCNB1, CCNB2, CDC25A, CDC25C, CDK1, MAD2L1, PLK1
hsa04218	Cellular senescence	ATM, ATR, CCNA2, CCNB1, CCNB2, CCNE1, CCNE2, CDC25A, CDK1, CHEK1, CHEK2, E2F1, FOXM1, MYBL2, TP53
hsa04115	p53 signaling pathway	ATM, ATR, CCNB1, CCNB2, CDK1, CDKN1A, CHEK1, CHEK2, GTSE1, MDM2, TP53
hsa04120	Ubiquitin mediated proteolysis	BRCA1, CDC20, UBE2C, UBE2S
hsa04068	FoxO signaling pathway	ATM, CCNB1, CCNB2, CDKN1B, PLK1, PLK4
hsa03030	DNA replication	FEN1, RNASEH2A
hsa03440	Homologous recombination	ATM, BRCA1, BRCA2, RAD54L
hsa01522	Endocrine resistance	E2F1, TP53
hsa03460	Fanconi anemia pathway	ATR, BRCA1, BRCA2, UBE2T
hsa05200	Pathways in cancer	BRCA2, CKS2, E2F1, TP53
hsa04151	PI3K-Akt signaling pathway	BRCA1, TP53
hsa05202	Transcriptional misregulation in cancer	ATM, TP53
hsa03410	Base excision repair	FEN1
hsa05222	Small cell lung cancer	CKS2, E2F1, TP53
hsa05215	Prostate cancer	E2F1, TP53
hsa05226	Gastric cancer	E2F1, TP53
hsa05220	Chronic myeloid leukemia	E2F1, TP53
hsa05219	Bladder cancer	E2F1, TP53
hsa05214	Glioma	E2F1, TP53
hsa05218	Melanoma	E2F1, TP53
hsa05212	Pancreatic cancer	BRCA2, E2F1, TP53
hsa05224	Breast cancer	BRCA1, BRCA2, E2F1, TP53
hsa05223	Non-small cell lung cancer	E2F1, TP53

3. Discussion

UBE2C is a member of the E2 ubiquitin-conjugating enzyme family, it plays a key role in the ubiquitination system in cooperation with APC/C. It is involved in mitotic cyclin B degradation, promoting the transition from the M phase to the G1 phase of the cell cycle [28,29]. Therefore, it is likely that aberrant *UBE2C* overexpression, leading to changes in ubiquitination, might be involved in uncontrolled cell proliferation, which is one of the main features of cancers.

The current study shows, for the first time, a global analysis of *UBE2C* expression in a wide array of tumors. Our results demonstrated that *UBE2C* is upregulated in all 27 different cancers examined and this is in agreement with previous reports showing the increased somatic expression of *UBE2C* in various tumor types. Indeed, the overexpression of *UBE2C* was previously reported in hepatocellular carcinoma, thyroid, colon, breast, and lung cancer [8–13] and our data confirmed this upregulation (Figure 1). To give an example, in a study conducted by Qin et al., the overexpression of *UBE2C* was found in breast cancer, including the basal-like (BL) subtype [30], and our study also showed this significant upregulation in both BL1 and BL2. Importantly, our investigation showed

that overexpression of *UBE2C* is a common feature of all 27 human cancers tested in this study, suggesting it acts as a proto-oncogene. Moreover, the current study revealed the significant *UBE2C* overexpression across the histological and molecular subtypes of different tumors mentioned in the results and also possible associations between this expression and different patient conditions, such as drinking and smoking habits in PAAD and LUAD, respectively. *UBE2C* expression, based on patients' pathological stages, showed that *UBE2C* overexpression can be involved in tumor progression and invasion. Our data broaden the observations of previous reports [15,30,31].

We showed that patients with higher expression of *UBE2C* had a shorter overall survival time and worse prognosis, and *UBE2C* higher-expression levels also resulted in worse DFS prognosis in many cancers, confirming that *UBE2C* overexpression results in poor clinical outcomes in many tumors. As an example, Qin et al. found that the *UBE2C* upregulation was associated with poor prognosis in breast cancer [30], and our data also revealed this issue as shown in Figure 3.

Furthermore, our study identified coexpression genes associated with the *UBE2C* protein network. Those genes with strong and very strong positive correlations with *UBE2C* expression in all cancers (Table 2) are involved in the cell cycle process. These include members of the kinesin family (*KIF20A*, *KIF18B*, *KIFC1*, and *KIF4A*), with roles in mitotic spindle maintenance, chromosome segregation, and microtubule depolarization [32,33]; *AURKB* as a regulator of chromosome segregation during mitosis [34]; *TROAP* (a member of the cell adhesion molecule complex) with a role in centrosome integrity during cell cycle progression [35]; *TPX2* (a microtubule-associated protein) having a key function in mitotic spindle formation [36]; *PLK1* as a regulator of cell division and maintenance of genome stability and spindle assembly, and DNA damage response [37,38]; *CDK1* (a cyclin-dependent kinase), involved in cell division with important functions in mitosis and driving cells into the S phase [39]; *CENPA* (a key component of the inner kinetochore plate), playing a role in chromosome segregation during oocyte meiosis [40]; *CDC20* as an activator of APC/C during mitosis for mitosis progression [41]; *MYBL2*, involved in cell cycle progression, cell survival, and cell differentiation [42]; *BUB1B* as a vital component of the mitotic checkpoint complex [43]; *CCNB1*, having an essential role in the transition of the cell cycle from G2 phase to mitosis [44,45]; *NCAPG* (a mitosis-related chromosome condensation protein) involved in the condensin I complex [46]; *SKA3*, controlling the correct exit from meiosis, migration of meiotic spindle, and stability of anaphase spindle [47]. Our data highlight the *UBE2C* network as one of the major protein networks involved in cancer and further investigation on their function in tumors might shed light onto new therapeutic strategies for cancer.

Interestingly, we also highlighted as *UBE2C* coexpressed genes the transcription factors, *FOXM1*, *E2F1*, and *RAD51*, which have their binding sites on the *UBE2C* promoter and enhancer regions, suggesting they are transcriptional regulators of this gene. Our results are in agreement with a previous report of a positive association of *FOXM1* with *UBE2C* in normal tissues and in tumors [48]. The identification of *RAD51*, which has a key role in the homologous recombination and repair of DNA [49], and mainly *E2F1* could be particularly relevant since *E2F1* contributes to the activation of genes involved in G1/S progression [50–53].

We also found a significant correlation between *UBE2C* and other important regulators of the cell cycle, including *CDC20*, which was previously identified as being co-expressed with *UBE2C* [54]. These data suggest that other genes identified in this study as part of the *UBE2C* protein network might play a role in many tumors. In particular, our data highlighted *MTFR2*, *MND1*, *FAM72D*, and *POC1A* as genes whose expression correlates with worse OS prognosis, suggesting their possible involvement in tumor progression and invasion.

In conclusion, the current study showed that *UBE2C* can be considered as a general tumor marker and study of its related pathways can help to discover common therapeutic targets for cancers. However, further functional studies are required to clarify the role of *UBE2C* in cancers.

4. Materials and Methods

In the current study, our investigations were performed using different bioinformatics tools and databases, including GEPIA [55] (a webserver which extracts data from the Cancer Genome Atlas (TCGA) data portal and the GTEx database of normal tissues. <http://gepia.cancer-pku.cn>), UALCAN (an interactive web portal for the in-depth analysis of TCGA gene expression data, <http://ualcan.path.uab.edu>) [56], and STRING databases (functional protein association networks (<https://string-db.org/>) [57]).

Here, we report the following investigations: *UBE2C* expression levels across all cancers and their subtypes, its differential gene expression analysis at different pathological stages, correlation between its expression and cancer prognosis in cancers, overall survival (OS) and the disease-free survival (DFS) analysis on the basis of *UBE2C* gene expression, investigation of genes with similar *UBE2C* expression patterns, and their associations in the *UBE2C* protein network.

To investigate *UBE2C* expression across 27 human tumor types compared to normal matches, we used the GEPIA webserver. One advantage of GEPIA is that it also uses normal data from the GTEx project to provide a reliable baseline for comparison. In most cancer research, normal tissues are prepared from areas adjacent to tumors, but they may be pre-cancerous tissue and not truly normal, healthy tissue. *UBE2C* expression between tumors, their matched normal, and data from the GTEx database in 27 tumor types were compared. These tumors included ACC, BLCA, BRCA, COAD, DLBC, ESCA, GBM, HNSC, KICH, KIRC, KIRP, LAML, LGG, LIHC, LUAD, LUSC, OV, PAAD, PRAD, READ, SKCM, STAD, TGCT, THCA, THYM, UCEC, and UCS. Six tumor types were excluded from the study (where differential expression was considered) due to their small numbers (13 and lower) or lack of normal samples, including cholangio carcinoma (CHOL), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), mesothelioma (MESO), pheochromocytoma and paraganglioma (PCPG), sarcoma (SARC), and uveal melanoma (UVM). Regarding parameter options, we used the ANOVA statistical method for differential gene expression analysis, selected $\log_2(\text{TPM} + 1)$ transformed expression data for plotting, TCGA tumors compared to TCGA normal and GTEx normal for matched normal data in plotting, $|\log_2\text{FC}|$ cutoff of 1, and a q-value cutoff of 0.01. Also, for cancers with different subtypes and conditions, we analyzed them using the UALCAN webserver.

To provide *UBE2C* expression box plots on the basis of patients' pathological stage (stage I, stage II, stage III, and stage IV group) in TCGA cancer types, we used the UALCAN webserver to get data from TCGA. In this analysis, cancers without normal matches or below two numbers in each stage were excluded from analysis.

Overall survival (OS) and the disease-free survival (DFS) analysis were also performed on the basis of *UBE2C* gene expression. Regarding hypothesis test, the GEPIA considers the Log-rank test. For this, we selected a hazards ratio (HR) based on the Cox PH model and also the 95% confidence interval information to show the 95% confidence interval (CI) as the dotted line. The *UBE2C* expression threshold of 50% (median value) was determined to split the *UBE2C* high-expression and low-expression cohorts. Therefore, samples with *UBE2C* expression levels higher and lower than 50% were applied as the high-expression cohort (cutoff-high) and the low-expression cohort (cutoff-low), respectively.

Correlation analysis between *UBE2C* and other genes was performed by pair-wise gene expression correlation analysis with the expression data of TCGA and GTEx, using the method of the Pearson correlation coefficient. At first, we searched for moderate, strong, and very strong *UBE2C* expression correlations (the Pearson correlation coefficient between 0.4 and 1) with other genes on average in all 27 cancers. Then, we investigated the *UBE2C* expression correlations with each gene individually in each cancer to see the exact correlation. We considered the following correlation coefficients: 0.00–0.19 as very weak, 0.20–0.39 as weak, 0.40–0.59 as fairly strong (also called moderate), 0.60–0.79 as strong, and 0.80–1.0 as very strong.

To provide the *UBE2C* protein network, the STRING database was used and most genes with medium to very strong correlations with *UBE2C* (extracted from TCGA cancer types using GEPIA) and also some important tumor suppressor proteins, such as ATR, ATM, BUB1B, BRCA1/2, CHK2, and CYLD, were searched in STRING.

To predict the RNA–RNA association and RNA–protein interaction between TSPYL and UBE2C, we used the RNAup webserver (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAup.cgi>) and RPISeq (<http://pridb.gdcb.iastate.edu/RPISeq/>), respectively. Also, to search for a D-box (one of the recognition amino acid sequences to be identified by APC/C in the ubiquitin–proteasome pathway) in TSPYL2 protein, we used the GPS-ARM tool (<http://arm.biocuckoo.org/down.php>).

Supplementary Materials: The following are available online at <http://www.mdpi.com/1422-0067/20/9/2228/s1>.

Author Contributions: H.D.: Conceptualization, data analysis, writing, D.D.: Technical support, M.C.: Data supervision, editing manuscript; S.O.: Supervision of the project, review, and editing the manuscript.

Funding: AIRC IG 2017 Id.20240 and IIGM institutional funds to SO and AIRC MGAF Id.20566 to MC.

Acknowledgments: We thank all members of the laboratory for useful discussions.

Conflicts of Interest: The authors declare no conflict of interest.

Truncating Mutations in UBAP1 Cause Hereditary Spastic Paraplegia

Mohammad Ali Farazi Fard,^{1,20} Adriana P. Rebelo,^{2,20} Elena Buglo,² Hamid Nemati,⁶ Hassan Dastsooz,^{1,3} Ina Gehweiler,^{4,5} Selina Reich,^{4,5} Jennifer Reichbauer,^{4,5} Beatriz Quintáns,¹² Andrés Ordóñez-Ugalde,¹² Andrea Cortese,² Steve Courel,² Lisa Abreu,² Eric Powell,⁷ Matt Danzi,² Nicole B. Martuscelli,¹⁵ Dana M. Bis-Brewer,² Feifei Tao,² Fariba Zarei,⁶ Parham Habibzadeh,^{1,8} Majid Yavarian,¹ Farzaneh Modarresi,⁹ Mohammad Silawi,¹ Zahra Tabatabaei,¹ Masoume Yousefi,¹ Hamid Reza Farpour,⁶ Christoph Kessler,^{4,5} Elisabeth Mangold,¹⁶ Xenia Kobeleva,¹⁷ Amelie J. Mueller,^{18,19} Tobias B. Haack,^{18,19} Mark Tarnopolsky,¹⁰ Ziv Gan-Or,¹¹ Guy A. Rouleau,¹¹ Matthis Synofzik,^{4,5} María-Jesús Sobrido,¹² Albena Jordanova,^{13,14} Rebecca Schüle,^{4,5,20} Stephan Zuchner,^{2,20} and Mohammad Ali Faghghi^{1,9,20,*}

The diagnostic gap for rare neurodegenerative diseases is still considerable, despite continuous advances in gene identification. Many novel Mendelian genes have only been identified in a few families worldwide. Here we report the identification of an autosomal-dominant gene for hereditary spastic paraplegia (HSP) in 10 families that are of diverse geographic origin and whose affected members all carry unique truncating changes in a circumscribed region of *UBAP1* (ubiquitin-associated protein 1). HSP is a neurodegenerative disease characterized by progressive lower-limb spasticity and weakness, as well as frequent bladder dysfunction. At least 40% of affected persons are currently undiagnosed after exome sequencing. We identified pathological truncating variants in *UBAP1* in affected persons from Iran, USA, Germany, Canada, Spain, and Bulgarian Roma. The genetic support ranges from linkage in the largest family (LOD = 8.3) to three confirmed *de novo* mutations. We show that mRNA in the fibroblasts of affected individuals escapes nonsense-mediated decay and thus leads to the expression of truncated proteins; in addition, concentrations of the full-length protein are reduced in comparison to those in controls. This suggests either a dominant-negative effect or haploinsufficiency. *UBAP1* links endosomal trafficking to the ubiquitination machinery pathways that have been previously implicated in HSPs, and *UBAP1* provides a bridge toward a more unified pathophysiology.

Hereditary spastic paraplegia (HSP) represents a group of genetically highly heterogeneous rare inherited neurodegenerative diseases, which are characterized by the pathological hallmark of a length-dependent degeneration of corticospinal-tract axons (see GeneReviews in [Web Resources](#)).¹ Clinically, HSPs are marked by progressive spastic paraparesis, although the clinical presentation encompasses a wide spectrum of phenotypes. In pure forms of HSP, progressive spasticity and weakness in the lower extremities are the main features. In complex forms of HSP, additional clinical symptoms include cataracts, ataxia, epilepsy, cognitive impairment, peripheral neuropathy, optic neuropathy, and deafness (see GeneReviews in [Web Resources](#)).¹ The prevalence of HSP has been estimated to be 1.3–9.6 in 100,000 (see GeneReviews in [Web](#)

[Resources](#)).^{1,2} Thus far, at least 58 genes have been reported to cause HSP in a Mendelian fashion.³ Yet approximately 40% of affected persons are still not diagnosed even after whole-exome sequencing (WES). Furthermore, many of the genes reported in recent years have only been described in a few families.⁴

In an effort to further close this diagnostic gap in HSP, we have gathered a highly diverse sample of 10 families from six countries (Iran, 1; USA, 1; Germany, 4; Canada, 1; Bulgaria, 2; and Spain, 1). Prior to the initiation of this study, all participating affected individuals gave informed consent in agreement with each institutional review board. In one family of Persian origin, family 1, we were able to genetically ascertain a total of 14 affected individuals from three generations ([Figure 1](#)). Sequencing of an HSP gene panel and CNV

¹Persian BayanGene Research and Training Center, Shiraz, Iran; ²John P. Hussman Institute for Human Genomics, Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL 33136, USA; ³Italian Institute for Genomic Medicine, University of Turin, Turin 10126 Italy; ⁴Department of Neurodegenerative Diseases, Hertie-Institute for Clinical Brain Research and Center of Neurology, University of Tübingen, Tübingen 72076, Germany; ⁵German Center for Neurodegenerative Diseases, Tübingen 72076, Germany; ⁶Shiraz University of Medical Sciences, Shiraz, Iran; ⁷The Genesis Project foundation Miami, FL 33136, USA; ⁸Student Research Committee, Shiraz University of Medical Sciences, Shiraz, Iran; ⁹Center for Therapeutic Innovation and Department of Psychiatry and Behavioral Sciences, University of Miami, Miami, FL 33136 USA; ¹⁰Department of Pediatrics, McMaster University, Hamilton, Ontario L8S 4L8, Canada; ¹¹Department of Human Genetics, McGill University, Montréal, Québec H3A 0G4, Canada; ¹²Neurogenetics Group Instituto de Investigación Sanitaria, Hospital Clínico de Santiago, Santiago de Compostela 15706, Spain; ¹³Molecular Neurogenetics Group VIB-UAntwerp, Center for Molecular Neurology, University of Antwerp, Antwerpen 2000, Belgium; ¹⁴Molecular Medicine Center Department of Medical Chemistry and Biochemistry, Medical University, Sofia, Sofia 1431, Bulgaria; ¹⁵Department of Biology University of Miami, Miami, FL 33136, USA; ¹⁶Institute of Human Genetics University of Bonn, Bonn 53113, Germany; ¹⁷Department of Neurology, University of Bonn, Bonn 53113, Germany; ¹⁸Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen 72076, Germany; ¹⁹Centre for Rare Diseases, University of Tübingen, Tübingen 72076, Germany

²⁰These authors contributed equally to this work

*Correspondence: mfaghghi@med.miami.edu
<https://doi.org/10.1016/j.ajhg.2019.03.001>

© 2019 This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



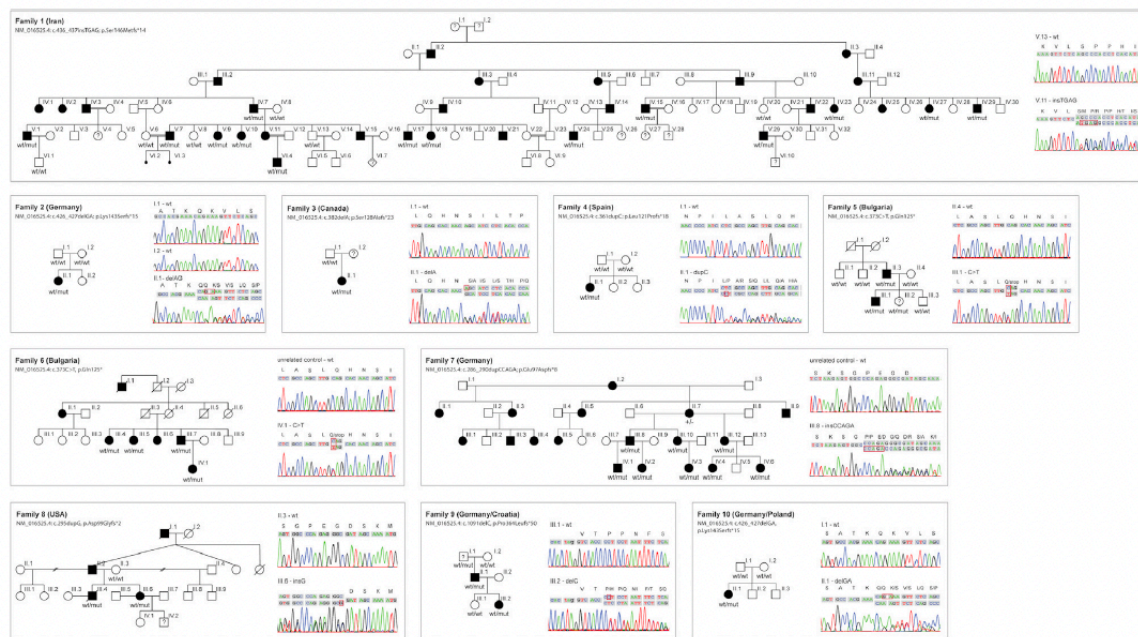


Figure 1. Pedigrees of HSP-Affected Families with UBAP1 Truncations

All pedigrees suggest an autosomal-dominant or a *de novo* Mendelian trait. HSP-affected individuals are marked by filled symbols; individuals with unclear affection status are marked by a question mark. “mut” depicts the presence of a causative allele. Sanger traces exemplify the confirmation of variants detected via next-generation sequencing. The penetrance of truncating UBAP1 variants is reduced: individual F5-III.2 was subjectively unaffected at age 14 but showed brisk reflexes of lower limbs, indicating potential dysfunction of the corticospinal tract. The 80-year-old grandfather of the index case in family 9 (F9-III.1) was unfortunately not available for a neurological examination but was reported to be in good health and without any indication of a gait disturbance.

analysis at the SPG4 locus were unremarkable. Subsequently, WES was performed in two affected individuals (V.1 and V.15). Bioinformatics analysis of the sequencing data used standard tools, including BWA aligner,⁵ FreeBayes,⁵ GATK,⁶ and GENESIS.⁷ Only non-synonymous variants with a minor-allele frequency of less than 0.0001 in gnomAD and in our in-house Iranian variant database (BayanGene; <http://www.bayangene.com>) of 1,500 exomes were further considered. Two heterozygous variants remained in *SVEP1* (chr9: 113137668; rs373655861; p.Thr3527Met hg19 [c.10580C>T]; GenBank: NM_153366.3) and *UBAP1* (chr9: 34241270; GenBank: NM_016525.4: c.436_437insTGAG [p.Ser146Metfs*14]), respectively. The *SVEP1* variant was ruled out by segregation studies involving Sanger sequencing of the entire pedigree. Thus, after confirmation of complete segregation, the truncating frameshift variant in ubiquitin-associated protein 1 (*UBAP1*) was considered as the causative allele in family 1 (Figure 1). This frameshift variant was not present in ExAC, gnomAD, GENESIS, nor in 1,500 Iranian genomes. It is predicted to truncate the protein at residue 158 out of 502 amino acids (GenBank: NM_016525.4). By including the Sanger-confirmed affected and unaffected participants (Figure 1), we performed parametric two-point linkage analysis by using the LINKAGE program, which rendered a two-point LOD score of 8.25 at the position of the *UBAP1* insertion.

We then searched the GENESIS database for additional families with *UBAP1* variants. GENESIS contains more than 3,000 exomes and genomes from affected persons with HSP and related disorders.⁷ We filtered for non-synonymous and truncating variants under an autosomal-dominant model with minor-allele frequency in gnomAD < 0.0001 and a minimum sequencing depth of 10 reads. We identified seven additional HSP families, all carrying truncating variants in *UBAP1* (Table 1). In addition, predictively truncating *UBAP1* variants were prioritized in two families (9 and 10) who underwent diagnostic exome sequencing at the University of Tuebingen. The detection of truncating alleles in all families is especially remarkable when one considers the almost complete constraint of *UBAP1* for loss-of-function (truncating) variation in the ~120,000 chromosomes in both ExAC and gnomAD, pLi = 0.95 and 0.92, respectively.⁸ We calculated the probability of significant enrichment of truncating variations in *UBAP1* in our HSP dataset compared to ExAC. In the GENESIS dataset we found seven such variants in a cohort of 567 HSP samples versus 0 truncating variants in 60,000 ExAC samples ($p = 6.187 \times 10^{-15}$ by Fisher test. Odds ratio = infinity). Five truncating variants were reported in the ~246,000 chromosomes in gnomAD, but none fell within the specific gene region containing the variants reported in this study.

Table 1. Detailed Genomic Locations of Detected Pathogenic Variants

Family ID	Genome Assembly (hg19)	Isoform 1 (GenBank: NM_016525.4) Expressed in Neurons		Isoform 4 (GenBank: NM_001171201.1) Canonical According to NCBI
		cDNA	Protein	Protein
1	chr9: 34241459–34241460	c.436_437insTGAG	p.Ser146Metfs*14	p.Ser210Metfs*14
2	chr9: 34241449–34241450	c.426_427delGA	p.Lys143Serfs*15	p.Lys207Serfs*15
3	chr9: 34241405–34241405	c.382del	p.Ser128Alafs*23	p.Ser192Alafs*23
4	chr9: 34241384–34241384	c.361dupC	p.Leu121Profs*18	p.Leu185Profs*18
5	chr9: 34241396–34241396	c.373C > T	p.Gln125*	p.Gln161*
6	chr9: 34241396–34241396	c.373C > T	p.Gln125*	p.Gln161*
7	chr9: 34241309–34241313	c.286_290dupCCAGA	p.Glu97Aspfs*8	p.Glu161Aspfs*8
8	chr9: 34241318–34241318	c.295dupG	p.Asp99Glyfs*2	p.Asp163Glyfs*2
9	chr9: 34249784–34249784	c.1091delC	p.Pro364Leufs*50	p.Pro428Leufs*50
10	chr9: 34241449–34241450	c.426_427delGA	p.Lys143Serfs*15	p.Lys207Serfs*15

We refer to isoform 1 throughout the text.

All additional variants and their segregation with disease in the additional families were confirmed by Sanger sequencing. On the basis of transcript GenBank: NM_016525.4, the identified variants were as follows (Table 1 and Figure 1): families 2 and 10 from Germany, c.426_427delGA (p.Lys143Serfs*15); family 3 from Canada, c.382del (p.Ser128Alafs*23); family 4 from Spain, c.361dupC (p.Leu121Profs*18); families 5 and 6 from Bulgaria (Roma ethnicity), c.373C>T (p.Gln125*); family 7 from Germany, c.286_290dupCCAGA (p.Glu97Aspfs*8); family 8 from the United States, c.295dupG (p.Asp99Glyfs*2), and family 9 from Germany c.1091delC (p.Pro364Leufs*50).

In families 2 and 4, a *de novo* occurrence of the truncating variant was confirmed (Figure 1). Families 5 and 6 were of self-declared Bulgarian Roma ethnicity and carried the same p.Gln125* variant, although the two index participants are from reportedly unrelated families. Evaluating the prevalence of this allele in the European Roma population and in Gypsy HSP-affected persons will require further studies.

The first manifesting symptom in all 30 UBAP1 mutation carriers from 10 families for whom detailed clinical data were available was a progressive spastic-gait disorder with a median age at onset of 8 years (interquartile range 4–9 years; oldest onset age 26 years; one asymptomatic mutation carrier (F5-III.2) aged 14 years (detailed clinical information in Table S1). At the time of examination (median disease duration 28 years; interquartile range 15–36 years), lower-limb spastic paraparesis was still the most prominent clinical feature in all affected mutation carriers; this was accompanied by brisk lower-limb tendon reflexes (all carriers, including asymptomatic carrier F5-III.2) and extensor plantar response in all but the youngest affected individual (F7-IV.6). Although brisk tendon reflexes of the upper limbs were frequently present (26 of 30; 87%) significant upper-limb spasticity was seen only in a single case (1/30; 3%; F4-II.1), consistent with a length-depen-

dent axonopathy of the corticospinal tract. Urinary urgency was reported in some cases (11 of 30; 37%), sensory deficits were absent or mild, and there was no evidence of peripheral neuropathy. In the majority of families (8/10; 80%), no additional signs or symptoms indicating affection of neuronal systems other than the corticospinal tract were seen, and the disease was accordingly classified as pure HSP. In family 7, however, seven out of nine family members had features of cerebellar involvement (such features included saccadic pursuit, gaze-evoked nystagmus, dysmetric saccades, and limb ataxia), features also present in family 9 (F9-II.1), indicating that the cerebellum is vulnerable to UBAP1 dysfunction at least in some cases.

Overall, truncating UBAP1 mutations are associated with a predominantly pure early-onset HSP phenotype; cerebellar involvement seems to be clustered in families and was observed in 2/10 families. Although there is thus minimal variation in terms of system involvement across families carrying UBAP1 mutations, phenotypic variability exists regarding the progression rate; for example, the disease progressed rather rapidly and led to early wheelchair dependency in families 2, 6, and 7 but was almost non-progressive in family 9 (F9-II.1 is still able to run and walk unlimited distances after 38 years of disease duration). Both intrafamilial as well as interfamilial variability are common or even the norm in HSP.³ A complete understanding of the phenotypic spectrum associated with UBAP1 mutations will require careful clinical evaluation of additional families carrying UBAP1 mutations.

UBAP1 is a member of the endosomal sorting complex required for transport -1 (ESCRT-I) complex and a regulator of vesicular trafficking processes, binds to ubiquitinated cargo proteins, and is essential for sorting endocytic ubiquitinated cargos into multivesicular bodies (MVBs).⁹ It also plays an important role in proteasomal degradation of ubiquitinated cell-surface proteins, including EGFR (epidermal growth factor receptor) and BST2 (bone marrow

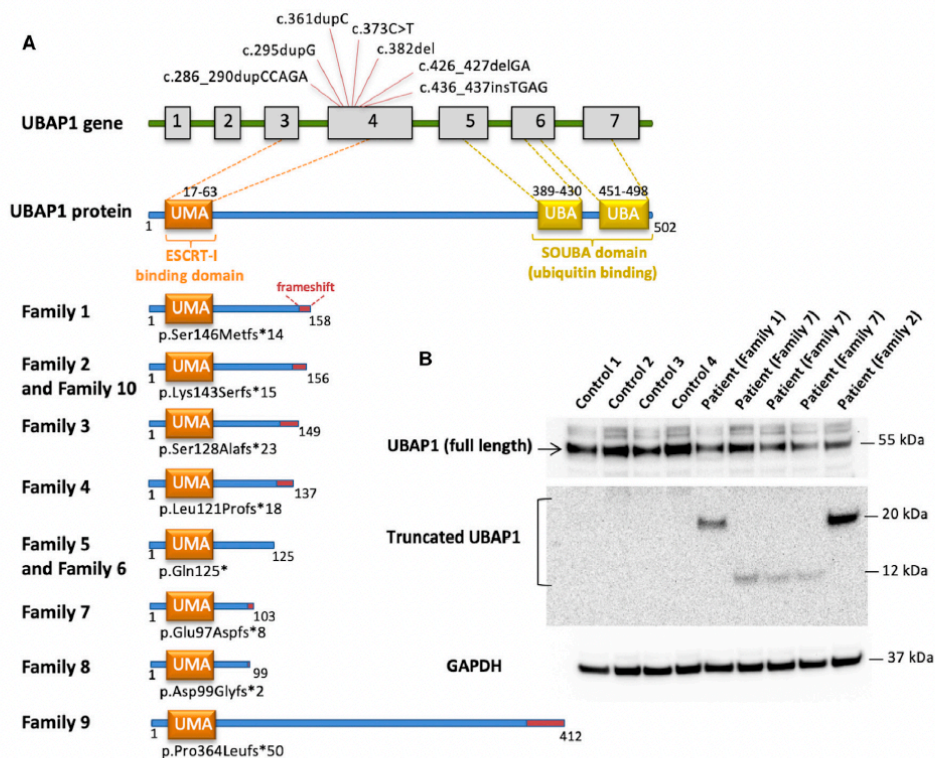


Figure 2. The Structure of *UBAP1* and Mutations Carried by Affected Individuals

(A) Schematic diagram showing all exons and UTRs of *UBAP1* on the basis of gene model GenBank: NM_016525.4. The gray boxes represent the coding sequence of *UBAP1*. All variants occurred in exon 4 of *UBAP1*. The UMA protein domain includes amino acids 17–63, and two UBAs include amino acids 389–430 and 451–498. All truncations are listed below; the preserving of the UMA domain is clearly depicted, but there is loss of the two SOUBA domains.

(B) Immunoblot analysis with an antibody recognizing amino acids 25–75 of *UBAP1* shows a notable decrease in the amount of full-length *UBAP1* in fibroblasts of the affected individuals from three different families when these cells are compared to control fibroblasts. Truncated *UBAP1* of the predicted sizes was detected in fibroblasts from affected individuals but not in control fibroblasts.

stromal cell antigen 2).⁹ *UBAP1* has two main domains: The UMA (*UBAP1*-MVB12-associated) domain in the N-terminal region (17–63 aa), which mediates the association with the ESCRT-I complex, and a SOUBA (solenoid of overlapping ubiquitin-associated domains) domain in the C-terminal region (389–498 aa).^{9,10} Both domains allow *UBAP1* to act as a molecular bridge connecting the endosomal trafficking pathways to the ubiquitination machinery. In an effort to decipher the pathophysiology of *UBAP1* in HSP, we noted that all but one of the identified changes fall within a circumscribed area of the protein between Asp 99 and Ser 146; the change in family 9 at Pro364 was the only outlier (GenBank: NM_016525.4). Interestingly, disease progression in this family has been dramatically slower than in the other families: the disease has been almost stationary over decades (see above), pointing toward a possible genotype-phenotype correlation. Yet, all changes preserve the UMA domain but cause a loss of the SOUBA domain.¹⁰ It has been shown that mutagenesis of the SOUBA domain in *UBAP1* strongly reduces its interac-

tion with ubiquitinated proteins (Figure 2).¹⁰ To determine whether the observed truncating variants would lead to nonsense-mediated mRNA decay and haploinsufficiency, we evaluated both the RNA and protein expression of mutant alleles. RT-PCR was performed on RNA extracted from the fibroblasts of an affected individual, and the RNA was sequenced by the Sanger method. Surprisingly, the c.436_437insTGAG was detected in the affected person's cDNA, indicating escape of nonsense-mediated mRNA decay (Figure S1). Next, we performed immunoblot analysis to evaluate both wild-type and potential truncated mutant *UBAP1*. Total protein extracts were probed with an antibody raised against the N-terminal region of *UBAP1* (amino acids 25–75), a part of the protein preserved in mutant *UBAP1* proteins. The protein levels measured in affected individuals were compared with those in four control fibroblasts and normalized to GAPDH levels. Immunoblots showed decreased protein levels of full-length *UBAP1* in fibroblasts from affected individuals; in addition, the truncated protein was detected (Figure 2). The reduced

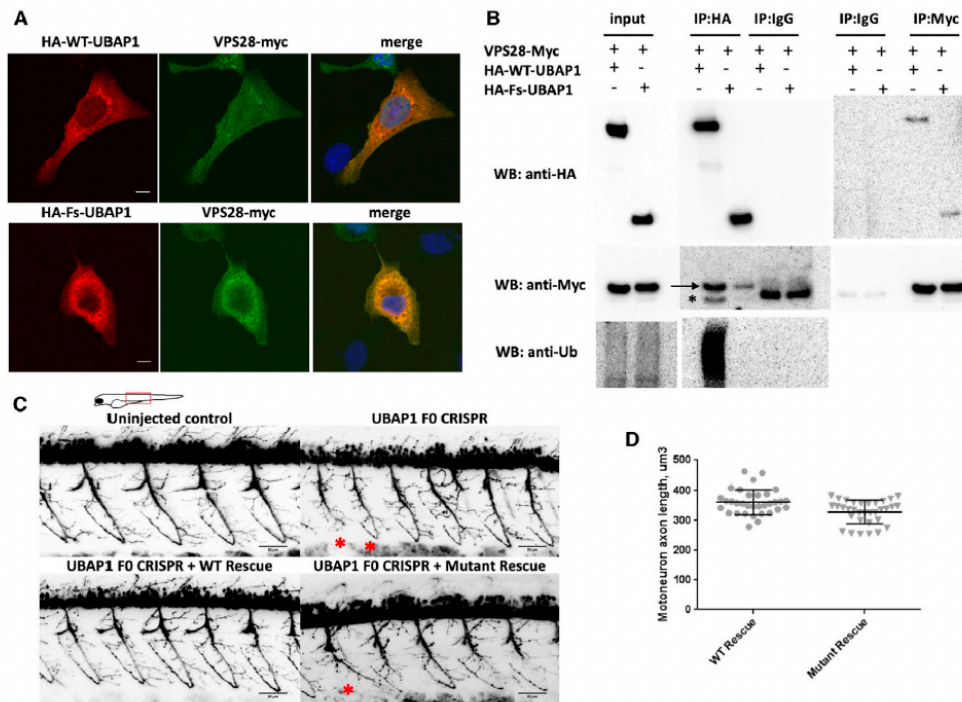


Figure 3. Functional *In Vitro* and *In Vivo* Studies of Truncated UBAP1

(A) Immunostaining of U2OS cells transfected with HA-WT-UBAP1 HA-Fs-UBAP1 (p.Leu121Profs*18) and VPS28-Myc. Both wild-type and mutant UBAP1 co-localize with VPS28-Myc.

(B) A co-immunoprecipitation assay shows protein-protein interaction between VPS28-Myc and both HA-WT-UBAP1 and HA-Fs-UBAP1. Ubiquitinated proteins co-immunoprecipitated with HA-WT-UBAP1 but not with HA-Fs-UBAP1. The arrow points to VPS28-Myc, and the asterisk below the arrow indicates the IgG band.

(C) Motor-neuron axons in Tg(*olig2::DsRed*) zebrafish embryos at 48 hpf. Embryos were injected with CRISPR Cas9 and sgRNAs against UBAP1; injection was supplemented with human RNA rescue of wild-type or truncated mutant UBAP1. Truncated and misshaped axons were more commonly observed with mutant hRNA rescue (indicated by asterisks). Scale bars represent 50 μm . The phenotypic difference between treated groups was evaluated by a Fisher exact test. Samples were assigned to either normal or affected categories on the basis of the presence of truncated and misshaped axons. The Fisher exact statistic value was determined to be 0.003; the result is significant at $p < 0.005$. Statistics describing normal versus affected phenotypes were calculated on the basis of the following sample sizes (number of embryos observed as having a phenotype). FO CRISPR + wild-type hRNA rescue: normal = 11, affected = 1. FO CRISPR + mutant hRNA rescue (family 4, p.Leu121Profs*18), normal = 9, affected = 15.

(D) Quantification of the individual motor-axon lengths. p values were calculated with a one-tailed Student's t test: $p = 0.0008$ and $n = 9$ (number of embryos in each experimental group; four axons were measured per embryo).

concentrations of the full-length protein in fibroblasts from affected individuals compared to controls along with the presence of the truncated protein could potentially lead to haploinsufficiency and/or a dominant-negative effect. To evaluate the effects of the truncated protein, we performed site-direct mutagenesis and generated a plasmid encoding the truncated protein fused to an HA tag at the N-terminal region. U2OS cells were co-transfected with either wild-type (HA-WT-UBAP1) or a truncated mutant (HA-Fs-UBAP1; p.Leu121Profs*18) together with its known binding partner VPS28-Myc. Both the wild-type and the truncated mutant co-localize with VPS28 (Figure 3A). This suggests that interaction with the ESCRT-I complex is preserved; however, the lack of the SOUBA domain, essential for ubiquitin binding, would be detrimental. Interestingly, overexpression of truncated

protein containing the UMA domain has been shown to result in a dominant-negative effect by inhibiting HIV-1 budding.¹⁰ It is thus possible that expression of the truncated protein in affected persons could cause a dominant-negative effect due to arrest of the ESCRT-complex without acquiring the ubiquitinated protein cargo.

We performed a co-immunoprecipitation (co-IP) assay to confirm the interaction between HA-Fs-UBAP1 and VPS28-Myc. HEK293T cells were co-transfected with VPS28-Myc and with either HA-WT-UBAP1 or HA-Fs-UBAP1 and immunoprecipitated with an anti-HA or an anti-Myc antibody and analyzed by immunoblot. Our results show that both wild-type and truncated UBAP1 co-immunoprecipitated with VPS28, confirming protein-protein interaction (Figure 3B). However, ubiquitinated proteins were co-immunoprecipitated with the HA-WT-UBAP1 but not

with HA-Fs-UBAP1. It has previously been shown that siRNA depletion of UBAP1 in HeLa cells causes clustering of early-endosome accumulation of ubiquitinated proteins and enlargement and clustering of LAMP1-positive late endosomes and lysosomes.⁹ In fibroblasts of affected persons carrying UBAP1 mutations, however, none of these changes could be observed (Figure S2), even after exposure of cells to stress conditions. It therefore appears unlikely that loss of one *UBAP1* allele results in the gross failure of multivesicular body sorting.

To investigate the effects of the truncated protein *in vivo*, we generated a zebrafish model with *UBAP1* knockdown. We used a transgenic fish with fluorescently labeled motoneuron Tg(*olig2::DsRed*).¹¹ Embryos were injected with CRISPR Cas9 and sgRNAs against *UBAP1* supplemented with human RNA rescue of either wild-type or truncated mutant UBAP1. At 48 hours post-fertilization (hpf), embryos were imaged *in vivo* with a confocal microscope. We observed significantly more truncated and misshaped axons in the mutant rescued embryos than in the wild-type rescued embryos (Figure 3C). Motoneuron axon lengths in the truncated mutant were significantly ($p = 0.0008$) shorter than those of the wild-type (Figure 3D). This result supports the pathogenic effects of the truncated protein *in vivo*.

In summary, we present strong genetic evidence that truncating mutations in *UBAP1* cause a relatively frequent form of HSP. *UBAP1* mutations were identified in a large Iranian kindred as well as in nine additional families with different ancestral backgrounds, including Bulgarian Roma, North American of European descent, German, Spanish, and Quebecois. All available affected persons in these families carried the respective mutation in *UBAP1*, although *UBAP1* has a strong loss-of-function constraint in the 60,000 individuals studied in the ExAC dataset. In two families we were also able to show a *de novo* occurrence of the variants. In our dataset of 567 families affected by dominant HSP, *UBAP1* accounts for 1.2% of cases. Evaluating the full allelic and clinical spectrum in this gene will require further studies. Because *UBAP1* links two cellular pathways previously involved in HSP, this finding consolidates our current understanding of the pathophysiology of HSP and points to potential novel drug targets.

Supplemental Data

Supplemental Data can be found online at <https://doi.org/10.1016/j.ajhg.2019.03.001>.

Acknowledgments

We are thankful to the families who participated in this research. We thank Els De Vriendt for the excellent technical assistance. S.Z. and R.S. are supported by National Institutes of Health grant R01NS072248. The project was further supported by the European Union's Horizon 2020 research and innovation program under grant 779257 (Solve-RD; R.S.) and in the context of the ERA-NET Cofund action N° 643578 by the German Bundesministerium für Bildung und Forschung (BMBF) (01GM1607) for the E-Rare-3

network PREPARE (M.S., S.R., and associated partner S.Z.). M.J.S. was supported by Instituto de Salud Carlos III grant PS09/01830 and FEDER (Fonds Européen de Développement Économique et Régional). A.J. is supported by an Excellence (TOP) grant of the University of Antwerp. The study was partly supported by NIMAD research grant (940714) awarded to MAF. The views expressed are those of the author(s) and not necessarily those of the National Health Service, the National Institute for Health Research, or the Department of Health. T.B.H. was supported by the BMBF through the "Juniorverbund in der Systemmedizin" "mitOmics" (FKZ 01ZX1405C to T.B.H.).

Declaration of Interests

The authors declare no competing interests.

Received: December 9, 2018

Accepted: February 27, 2019

Published: March 28, 2019

Web Resources

OMIM, www.omim.org

GENESIS platform, <https://www.tgp-foundation.org/>

GeneReviews, Fink, J.K. (1993). Hereditary Spastic Paraplegia Overview, <https://www.ncbi.nlm.nih.gov/books/NBK1509/>

References

- Fink, J.K. (2013). Hereditary spastic paraplegia: Clinico-pathologic features and emerging molecular mechanisms. *Acta Neuropathol.* 126, 307–328.
- Parodi, L., Fenu, S., Stevanin, G., and Durr, A. (2017). Hereditary spastic paraplegia: More than an upper motor neuron disease. *Rev. Neurol. (Paris)* 173, 352–360.
- Schüle, R., Wiethoff, S., Martus, P., Karle, K.N., Otto, S., Klebe, S., Klimpe, S., Gallenmüller, C., Kurzwelly, D., Henkel, D., et al. (2016). Hereditary spastic paraplegia: Clinicogenetic lessons from 608 patients. *Ann. Neurol.* 79, 646–658.
- Bis-Brewer, D.M., and Züchner, S. (2018). Perspectives on the genomics of HSP beyond Mendelian inheritance. *Front. Neurol.* 9, 958.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303.
- Gonzalez, M., Falk, M.J., Gai, X., Postrel, R., Schüle, R., and Züchner, S. (2015). Innovative genomic collaboration using the GENESIS (GEM.app) platform. *Hum. Mutat.* 36, 950–956.
- Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291.
- Stefani, F., Zhang, L., Taylor, S., Donovan, J., Rollinson, S., Doyotte, A., Brownhill, K., Bennion, J., Pickering-Brown, S., and Woodman, P. (2011). UBAP1 is a component of an endosome-specific ESCRT-I complex that is essential for MVB sorting. *Curr. Biol.* 21, 1245–1250.
- Agromayor, M., Soler, N., Caballe, A., Kueck, T., Freund, S.M., Allen, M.D., Bycroft, M., Perisic, O., Ye, Y., McDonald, B., et al. (2012). The UBAP1 subunit of ESCRT-I interacts with ubiquitin via a SOUBA domain. *Structure* 20, 414–428.
- Kucenas, S., Snell, H., and Appel, B. (2008). *nkx2.2a* promotes specification and differentiation of a myelinating subset of oligodendrocyte lineage cells in zebrafish. *Neuron Glia Biol.* 4, 71–81.



LncRNA–miRNA–mRNA Networks of Gastrointestinal Cancers Representing Common and Specific LncRNAs and mRNAs

Hassan Dastsooz^{1,2,3†}, Ahad Alizadeh^{4†}, Parham Habibzadeh⁵, Ali Nariman⁶, Asieh Hosseini⁷, Yaser Mansoori^{8,9} and Hamed Haghi-Aminjan^{10*}

¹Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy, ²Candiolo, C/o IRCCS, IIGM-Italian Institute for Genomic Medicine, Turin, Italy, ³Candiolo Cancer (IT), FPO-IRCCS, Candiolo Cancer Institute, Turin, Italy, ⁴Medical Microbiology Research Center, Qazvin University of Medical Sciences, Qazvin, Iran, ⁵Research Center for Health Sciences, Institute of Health, Shiraz University of Medical Sciences, Shiraz, Iran, ⁶Genetics and Molecular Biology Department, Isfahan University of Medical Sciences, Isfahan, Iran, ⁷Razi Drug Research Center, Iran University of Medical Sciences, Tehran, Iran, ⁸Department of Medical Genetics, Fasa University of Medical Sciences, Fasa, Iran, ⁹Noncommunicable Diseases Research Center, Fasa University of Medical Sciences, Fasa, Iran, ¹⁰Pharmaceutical Sciences Research Center, Ardabil University of Medical Sciences, Ardabil, Iran

OPEN ACCESS

Edited by:

Ramkrishna Mitra,
Thomas Jefferson University,
United States

Reviewed by:

Sayed Haider Abbas Faza,
Northwest A. and F. University, China
Lei Wang,
Changsha University, China

*Correspondence:

Hamed Haghi-Aminjan
hamedhaghi.a@gmail.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
RNA,
a section of the journal
Frontiers in Genetics

Received: 09 October 2021

Accepted: 19 November 2021

Published: 24 January 2022

Citation:

Dastsooz H, Alizadeh A, Habibzadeh P,
Nariman A, Hosseini A, Mansoori Y and
Haghi-Aminjan H (2022)
LncRNA–miRNA–mRNA Networks of
Gastrointestinal Cancers Representing
Common and Specific LncRNAs
and mRNAs.
Front. Genet. 12:791919.
doi: 10.3389/fgene.2021.791919

Gastrointestinal (GI) cancers are responsible for approximately half of cancer-related deaths, highlighting the need for the identification of distinct and common features in their clinicopathological characteristics. Long ncRNA (lncRNAs), which are involved in competitive endogenous RNA (ceRNA) networks with critical roles in biological processes, constitute a substantial number of non-coding RNAs. Therefore, our study aimed to investigate the similarities and differences in the ceRNA networks of The Cancer Genome Atlas (TCGA)-GI cancers. We performed a comprehensive bioinformatics analysis of ceRNA networks for TCGA-GI cancers in terms of the differential mRNA, lncRNA, and miRNA expression levels, ceRNA networks, overall survival analysis, correlation analysis, pathological cancer stages, and gene set enrichment analysis. Our study revealed several common and distinct mRNAs and lncRNAs with prognostic values in these networks. It was specifically noteworthy that *MAGI2-AS3* lncRNA was found to be shared in almost all GI cancers. Moreover, the most common shared mRNAs between GI cancers were *MEIS1*, *PPP1R3C*, *ADAMTSL3*, *RIPOR2*, and *MYLK*. For each cancer ceRNA network, we found that the expression level of a number of lncRNAs and mRNAs was specific. Furthermore, our study provided compelling evidence that several genes, most notably *KDEL1*, can act as novel proto-oncogenes in cancers. This, in turn, can highlight their role as new prognostic and therapeutic targets. Moreover, we found cell cycle and extracellular matrix structural constituent as the top shared KEGG and molecular function, respectively, among GI cancers. Our study revealed several known lncRNAs and known and unknown mRNAs in GI cancers with diagnostic and prognostic values.

Keywords: tumor biomarkers, gastrointestinal cancers, long-non-coding RNA, The Cancer Genome Atlas (TCGA), competitive endogenous RNA (ceRNA)

1 INTRODUCTION

The gastrointestinal (GI) system consists of the substantial cellular mass in the human body, and its cancers are among the most common malignancies in different populations, accounting for 35% of the total deaths due to cancers (Arnold et al., 2020). Over the past decade or two, tremendous efforts have been invested into the identification of the molecular and biological processes responsible for the development of these cancers, mainly for colorectal cancer, hepatic cancer, gastric cancer, and head and neck squamous cell carcinomas (Sharma et al., 2018). In particular, identification of distinct and common features in their molecular and biological processes and also their clinical presentations can help shed light on the development and identification of diagnostic and therapeutic biomarkers.

More than 90% of the mammalian genome is transcribed into non-coding RNA (ncRNA), with a considerable number of them consisting of long ncRNAs (lncRNAs), transcripts over 200 nucleotides (nt) long (Sulayman et al., 2019). Increasing evidence has shown that lncRNAs are involved in competitive endogenous RNA (ceRNA) networks playing critical biological functions including the regulation of major cellular processes, namely, proliferation, differentiation, apoptosis, and stress response (Niu et al., 2020b; Li et al., 2021). Therefore, our study aimed to investigate the similarities and differences in the ceRNA networks of TCGA-GI cancers with more focus on lncRNAs. We included GI cancers with 11 and more than 11 matched normal tissues. These cancers were as follows: colon adenocarcinoma (COAD), rectal adenocarcinoma (READ), esophageal carcinoma (ESCA), stomach adenocarcinoma (STAD), head and neck squamous cell carcinoma (HNSC), and liver hepatocellular carcinoma (LIHC). Since HNSC is involved in the mucosa of the aero-digestive tract and also occurs in the oral cavity and salivary glands, we considered this cancer in our study as well.

Here, we report different and common shared lncRNAs and mRNAs involved in the ceRNA networks of these cancers. Moreover, our study shows the common and specific pathways in these cancers. Our finding also shows that a combination of lncRNAs and mRNA with prognostic values can promote their use for diagnostic and therapeutic aims. Furthermore, we propose a number of mRNAs that can function as new proto-oncogenes or tumor suppressors in their corresponding cancers, highlighting their diagnostic and therapeutic aspects.

Since our study reveals common shared and distinct mRNAs, lncRNAs, miRNAs, KEGG, GO-MF, GO-BP, and GO-CC among GI cancers, the identified results can help other researchers to use the data for further functional studies and differential diagnostic and prognostic strategies.

2 MATERIALS AND METHODS

2.1 Data Collection and the Differential Expression Levels of lncRNAs, mRNAs, and miRNAs

The Cancer Genome Atlas (TCGA) (<https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>)

contains raw data of genomic, epigenomic, transcriptomic, and proteomic experiments from over 20,000 primary tumor tissues and their matched normal counterparts from 33 cancer types. In this study, we used an R/Bioconductor package, GDCRNATool developed by Ruidong Li et al. (Li et al., 2018) to download, organize and analyze lncRNAs, mRNAs, and miRNAs and clinical data of patients with GI cancers (HNSC, ESCA, STAD, LIHC, COAD, and READ) from TCGA.

Based on the following advantages, we used GDCRNATool for our analysis: One—GDCRNATools is an easy-to-use package for researchers with little coding experience. Two—It includes all data needed for performing the entire analysis smoothly in a very convenient way to download, organize, and perform comprehensive RNA expression analysis of TCGA data, with main focus on construction of the lncRNA–mRNA–miRNA-related ceRNA networks in cancer. Three—It gives not only genes involved in ceRNA networks but also all differentially expressed protein-coding genes, non-coding genes, and their GO terms and pathways. Therefore, it also has a flexibility to do other analyses that are not in the scopes of ceRNA network analysis. As an example, the differential expression data extracted from GDCRNATool can be used for considering the functional studies. Four—Moreover, to show that our differential expression data have an acceptable performance comparable to the state-of-the-art methods, we validated our expression data by looking for the expression of one gene (*UBE2C*) that was shown to be upregulated in all cancers (Dastsooz et al., 2019). Our differential expression data extracted from GDCRNATool also showed its higher expression in all these GI cancers. So, this tool can validate the differential expression data.

To validate our results, we also performed all analyses for the example that the developer of the tool did in their paper (for CHOL cancer). After confirmation of consistency between our results and theirs, we performed our analyses for GI cancers. So, we confirmed that this method of analysis is reproducible for all TCGA cancers. In our study, several analyses were carried out using GDCRNATools, which included differential gene expression analysis, univariate survival analysis, competing endogenous RNA network analysis, and functional enrichment analysis.

The appropriate data of these GI tumors and their matched adjacent non-tumor tissues were extracted. We started our analysis with the following sample numbers of RNA sequencing, miRNA, and clinical data for each cancer: COAD: 521 mRNA data, 465 miRNA data, and 459 clinical data; ESCA: 173 mRNA data, 200 miRNA data, and 185 clinical data; HNSC: 546 mRNA data, 569 miRNA data, and 528 clinical data; LIHC: 424 mRNA data, 425 miRNA data, and 377 clinical data; READ: 177 mRNA data, 165 miRNA data, and 171 clinical data; STAD: 407 mRNA data, 491 miRNA data, and 443 clinical data.

In this package, we downloaded RNA sequencing, mature miRNA, and clinical data from TCGA. Then, we parsed RNA sequencing metadata, filtered duplicated samples in this RNA metadata, and filtered non-primary tumor and non-solid tissue normal samples in them. Then, we parsed miRNAs metadata, filtered duplicated samples in these miRNAs data, and filtered non-primary tumor and non-solid tissue normal samples in the

miRNAs metadata. Next, we separately merged each RNA sequencing, miRNAs, and clinical data. After that, we normalized RNA sequencing data and miRNAs data (using `gdcVoomNormalization`). We then used the `gdcDEAnalysis` function to have a whole differential expression and the `gdcDEReport` function for differentiation expression of each miRNA, lncRNA, and mRNA. Using the `gdcCEAnalysis` function considering data of differential expression of lncRNA and protein-coding genes, and extracting data of Starbas tool, we could decipher miRNA–lncRNA and miRNA–mRNA interactions and ceRNA networks of each GI cancer.

We used this package to have the `gdcParseMetadata` function for the efficient organization of RNA and clinical data. With this tool, we applied the `gdcFilterDuplicate` function to remove duplicated samples and `gdcFilterSampleType` to filter out samples without primary tumors or matched normal counterparts. Raw counts data were normalized using the `gdcVoomNormalization` function, which applied the TMM method in `edgeR` and the voom method in `limma`. Low-expression genes with `logcpm` less than one were excluded from the analysis. In this R package, the `gdcDEAnalysis` function, which included `limma`, `edgeR`, and `DESeq2`, was used for the identification of differentially expressed genes (DEGs) and miRNAs between primary GI tumors and their matched normal tissues. Visualization methods were volcano, scatter, and bubble plots, and also three simple shiny apps were used in `GDCRNATools`. All the figures were plotted using the `ggplot2` package included in this tool.

2.2 Construction of the ceRNA Network

Using `GDCRNATool`, we investigated ceRNA networks in GI cancers. This tool with its `gdcCEAnalysis` function considers hypergeometric test (significantly commonly shared miRNAs between lncRNA and mRNA are detected), Pearson correlation analysis (lncRNA and mRNA with positive correlations are picked up), regulation similarity analysis (commonly shared miRNAs with similar function for regulation of the lncRNA and mRNA are considered), and sensitivity Pearson partial correlation to construct ceRNA networks. Using this tool, we first extracted several miRNAs that were correlated among both the lncRNAs and mRNAs. Then, this tool identified lncRNAs and mRNAs with positive expression correlation (Pearson correlation coefficient). Finally, we gained miRNAs with similar regulative effects on lncRNAs and mRNAs. In our analysis, we applied `lncTarget` (to get miRNA–lncRNA interactions) and `pcTarget` (to consider miRNA–mRNA interactions) data along with `gdcCEAnalysis` function to extract miRNA–target interactions predicted from several datasets. `GdcCEAnalysis` function considers data of miRNA–mRNA and miRNA–lncRNA interaction databases, for example, the interaction between miRNA and mRNA and also mRNA predicted by online datasets such as Starbas. In our analysis, lncRNA–miRNA–mRNA interactions extracted as edges and nodes were visualized in `Cytoscape 3.7.2`.

2.3 mRNA–lncRNA Correlation Analysis

To find the main mRNA in each ceRNA network of GI cancers, using `GDCRNA` package with `gdcCEAnalysis` function, we also investigated lncRNAs–mRNAs correlation.

2.4 Survival Analysis

We performed univariate survival analysis using the `gdcSurvivalAnalysis` function in the `GDCRNA` package with the selection of Kaplan–Meier (KM) analysis. In this analysis, the patients were divided into high- and low-expression groups according to the median.

2.5 Pathological Tumor Stages

We looked for expression of lncRNAs and mRNAs identified in ceRNA networks of GI cancers across their pathological tumor stages using GEPIA2 web server (<http://gepia2.cancer-pku.cn/#index>) (Tang et al., 2019).

2.6 Protein–Protein Interaction Network

We used the STRING database (<https://string-db.org>) (Szklarczyk et al., 2019) to analyze the possible protein–protein interactions between possible novel biomarkers identified in ceRNA networks of these cancers.

2.7 Functional Enrichment Analysis

Using `GDCRNATool`, we investigated the role of the genes that were found to be involved in the GI cancers, in biological processes (BP), molecular functions (MF), and cellular components (CC). Using `gdcEnrichAnalysis`, we also carried out Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG).

3 RESULTS

3.1 Differential Expression Profile of ceRNA Network Components in GI Cancers

In the current study, we found different numbers of up- and downregulated mRNAs, lncRNAs, and miRNAs in HNSC, ESCA, STAD, COAD, READ, and LIHC given in Table 1. Regarding COAD, these differentially expression numbers and the lncRNAs were also previously listed (Poursheikhani et al., 2020). Regarding STAD and LIHC, some differentially expression numbers have been reported by Zhang et al. (2020a) and Zheng and Yu (2021), respectively.

3.2 ceRNA Networks in GI Cancers

Our study revealed the following numbers of differentially expressed mRNAs (DEmRNAs), lncRNA (DElncRNA), and miRNAs (DemiRNAs) maintained in the ceRNA networks of each cancer (Supplementary Table S1): HNSC (6 lncRNAs, 30 miRNAs, and 64 mRNAs), ESCA (4 lncRNAs, 17 miRNAs, and 15 mRNAs), COAD [9 lncRNAs, 37 miRNAs, and 71 mRNAs, the ceRNA network for COAD cancer was previously shown in a different way (Poursheikhani et al., 2020)], STAD (4 lncRNAs, 18 miRNAs, and 29 mRNAs), READ (9 lncRNAs, 21 miRNAs, and 45 mRNAs), and LIHC [10 lncRNAs, 30 miRNAs, and 64 mRNAs, its ceRNA network was shown in a different figure by Zheng and Yu (2021)] (Supplementary Table S1).

Our findings showed that LIHC, COAD, and READ had more ceRNA networks. It is worth noting that our study revealed *MAGI2-AS3* lncRNA to be shared among almost all ceRNA networks of GI

TABLE 1 | Number of differentially expressed mRNAs, lncRNAs, and miRNAs in GI cancers.

Tumor type	COAD		READ		STAD		HNSC		LIHC		ESCA	
	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down
mRNA	1,094	1,901	1,169	1,790	935	1,192	940	1,103	715	1,448	610	722
lncRNA	128	77	181	53	119	51	76	32	68	80	49	49
miRNA	170	160	165	114	67	59	88	81	59	71	46	33

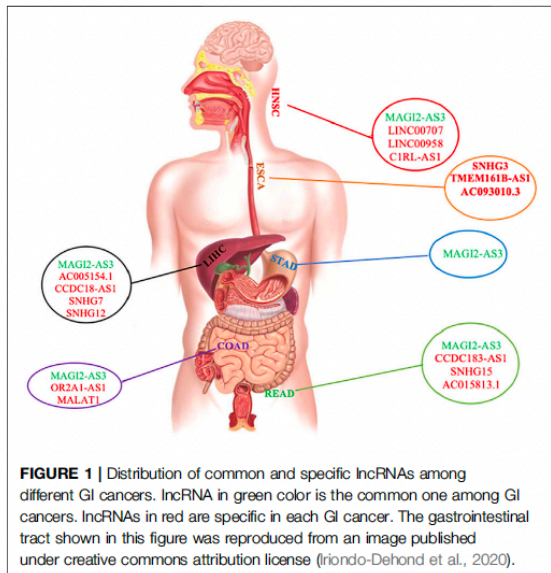


FIGURE 1 | Distribution of common and specific lncRNAs among different GI cancers. lncRNA in green color is the common one among GI cancers. lncRNAs in red are specific in each GI cancer. The gastrointestinal tract shown in this figure was reproduced from an image published under creative commons attribution license (friondo-Dehond et al., 2020).

cancers, except for ESCA, indicating the involvement of the same pathways in GI cancers (Supplementary Table S2, Supplementary Figure S1, Figure 1). Furthermore, *PVT1* was common between LIHC, ESCA, and STAD; *GAS5*, *SNHG1*, and *SNHG20* were common between LIHC, COAD, and READ; *KCNQ1OT1* was common between STAD, HNSC, COAD, and READ; *H19* was found to be common between LIHC, HNSC, and COAD; and finally, *MIR17HG* was common between STAD, COAD, and READ. We also investigated commonly shared lncRNAs between COAD and READ (CRCs) and found the presence of *MAGI2-AS3*, *GAS5*, *SNHG1*, *KCNQ1OT1*, *SNHG20*, and *MIR17HG* in both cancers. We were also able to distinguish COAD and READ using differential expression of *H19*, *OR2A1-AS1*, and *MALAT1* (specific for COAD) and *SNHG15*, *AC015813.1*, and *CCDC183-AS1* (specific for READ) (Supplementary Table S2 and Supplementary Figure S1).

For each cancer ceRNA network, we found that the expression level of a number of lncRNAs was specific to each cancer: *SNHG12*, *AC005154.1*, *CCDC18-AS1*, and *SNHG7* were specific for LIHC; *SNHG3*, *TMEM161B-AS1*, and *AC093010.3* were specific for ESCA; *SNHG15*, *AC015813.1*, and *CCDC183-AS1* were specific for READ; *C1RL-AS1*, *LINC00707*, and *LINC00958* were specific for HNSC; *OR2A1-AS1* and *MALAT1* were specific for COAD

(Figure 1 and Supplementary Table S2). There was no specific lncRNA for ceRNA network of STAD. However, in comparison with other GI cancers, it was possible to distinguish it from other cancers (Figure 1, Supplementary Table S2, and Supplementary Figure S1).

Our study revealed that ceRNA networks with involvements of the following lncRNAs consist of a number of important mRNAs, indicating the presence of important ceRNA networks in their corresponding cancers: *MAGI2-AS3*, *SNHG15*, *KCNQ1OT1*, *SNHG1*, and *MIR17HG* in READ; *MAGI2-AS3*, *KCNQ1OT1*, *H19*, *MIR17HG*, *SNHG1*, and *MALAT1* in COAD; *MAGI2-AS3*, *KCNQ1OT1*, and *MIR17HG* in STAD; *MAGI2-AS3*, *PVT1*, *AC005154.1*, *H19*, *CCDC18*, and *SNHG1* in LIHC; *MAGI2-AS3*, *KCNQ1OT1*, *H19*, *C1RL-AS1*, *LINC00707*, and *LINC00958* in HNSC; and finally *TMEM161B-AS1* and *AC093010.3* in ESCA (Supplementary Figure S1).

We mentioned that *MAGI2-AS3* is shared in approximately all GI cancers. Then, we looked for proteins involved in its network and its significant interaction with them. We found several important ones shown in Supplementary Table S3. In all *MAGI2-AS3* networks, it has interacted with the following miRNAs: *has-miR-374a-5p* and *has-miR-374b-5p*. Its main mRNA partners that have positive correlations are *MEIS1*, *PPP1R3C*, *ADAMTSL3*, *RIPOR2*, and *MYLK*. It seems that the function of *MAGI2-AS3* is mainly dependent on its interaction with these five positive correlated genes in COAD, READ, and STAD. However, in LIHC, it is dependent on *ADAMTSL3*, *RIPOR2*, and *MYLK*, and in HNSC, it is dependent on the interaction with *MEIS1* and *PPP1R3C* (Supplementary Table S3 and Supplementary Figure S1).

To find shared mRNAs among ceRNA networks of GI cancers, we compared the networks and identified four mRNAs, which were common in most of them as follows: *MEIS1* and *PPP1R3C* among CC-1 (COAD, READ, STAD, and HNSC), and *ADAMTSL3* and *MYLK* among CC-2 (COAD, READ, STAD, and LIHC) (Table 2 and Supplementary Figure S1). Moreover, we found several important shared mRNAs between ceRNA networks of COAD and READ (CRC, cancer cluster 3:CC-3) (Table 2), which included *APC*, *EDIL3*, *FGFR2*, *HOMER1*, *MIER3*, *PII5*, *RBM28*, *RRS1*, *SCML1*, *SGPPI*, *TOMM34*, *TRAF5*, *WNT5A*, and *ZNF655*. For other cancer clusters, other important shared mRNAs as shown in Table 2 were identified. These data indicated the common pathways for ceRNA networks of each cancer cluster. However, our data also showed specific protein-coding genes for each cancer presented in Supplementary Table S2, demonstrating specific ceRNA pathways for each cancer.

TABLE 2 | Commonly shared mRNAs between ceRNA networks of GI cancers.

CC-1	<i>MEIS1</i> and <i>PPP1R3C</i>
CC-2	<i>ADAMTSL3</i> and <i>MYLK</i>
CC-3	<i>APC</i> , <i>EDIL3</i> , <i>FGFR2</i> , <i>HOMER1</i> , <i>MIER3</i> , <i>PI15</i> , <i>RBM28</i> , <i>RRS1</i> , <i>SCML1</i> , <i>SGPP1</i> , <i>TOMM34</i> , <i>TRAF5</i> , <i>WNT5A</i> , and <i>ZNF655</i>
CC-4	<i>ATP6B2</i> , <i>L1CAM</i> , <i>LPAR1</i> , <i>NTN1</i> , <i>PCDH7</i> , <i>TTL7</i> , <i>UST</i> , and <i>VAMP2</i>
CC-5	<i>ANKRD13B</i> and <i>RIPOR2</i>
CC-6	<i>CDC7</i> , <i>SOX12</i> , and <i>SRPX</i>
CC-7	<i>BUB1</i> and <i>LMNB2</i>
CC-8	<i>COL5A2</i> , <i>DI O 2</i> , <i>KDELC1</i> , <i>SPARC</i> , and <i>STX1A</i>
CC-9	<i>NFIX</i> and <i>PDZD2</i>
CC-10	<i>SPRY2</i>
CC-11	<i>RPS6KA5</i> (ESCA and READ) and <i>DAAM2</i> (ESCA and COAD)

CC-1: mRNAs shared between ceRNA networks of COAD, READ, HNSC, and STAD; CC-2: mRNA shared between ceRNA networks of COAD, READ, LIHC, and STAD; CC-3: mRNAs shared between ceRNA networks of COAD, and READ (CRCs); CC-4: mRNAs shared between ceRNA networks of STAD, and CRCs; CC-5: mRNA shared between ceRNA networks of LIHC, and CRCs; CC-6: mRNAs shared between ceRNA networks of LIHC, and COAD; CC-7: mRNAs shared between ceRNA networks of LIHC, and READ; CC-8: mRNAs shared between ceRNA networks of HNSC, and COAD; CC-9: mRNAs shared between ceRNA networks of HNSC, and STAD; CC-10: mRNAs shared between ceRNA networks of HNSC, and LIHC; C11: mRNAs shared between ceRNA networks of ESCA, and other GI, cancers.

3.3 OS Analysis

In the GI-ceRNA networks, we identified that the overexpression of *LINC00958* and *LINC00707* in HNSC and overexpression of *SNHG20* and *SNHG12* in LIHC were correlated with reduced OS and worse prognosis. However, overexpression of *SNHG20* in READ and *PVT1* in STAD and low expression of *MAGI2-AS3* in STAD were correlated with good prognosis (Figure 2). Moreover, OS analysis for mRNAs involved in ceRNA networks of these cancers showed prognostic values of several mRNAs given in Table 3 and Supplementary Figure S2. Among them, we identified some new prognostic biomarkers (reduced OS with worse prognosis) that have not been reported or supported by functional studies in their corresponding tumors as follows: upregulation of *KDELC1* (*POGLUT2*) in HNSC, downregulation of *RIPOR2* and *SMOC1*, and upregulation of *TMEM164*, *B3GNT5*, *ZNF607*, and *ANKRD13B* in LIHC. These genes showed reduced OS with worse prognosis. The downregulated genes with reduced OS and worse prognosis may function as tumor suppressors, but those overexpressed mRNAs with reduced OS and worse prognosis can act as proto-oncogenes.

It worth noting that for LIHC, combinations of prognostic values of *SNHG20* lncRNA and *BUB1* mRNA, and also *SNHG12* lncRNA and *TMEM164* mRNA (associated with reduced OS with worse prognosis), could strengthen the predictive value of these markers for this cancer. Each of the lncRNA–mRNA combinations was in the same ceRNA network.

3.4 Pathological GI Tumor Stages

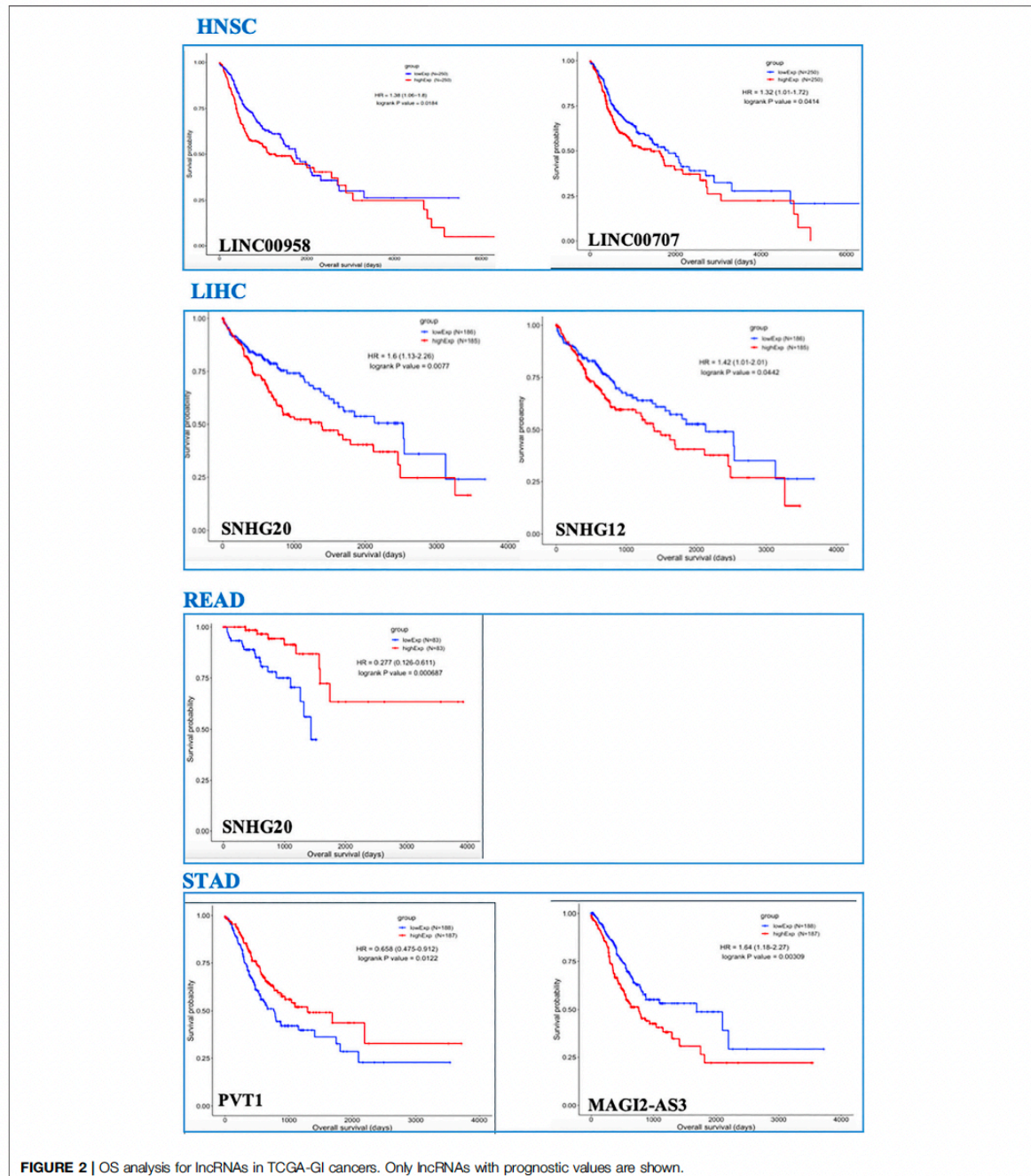
We found that the expression of some lncRNAs with worse prognosis in ceRNA networks of GI cancers had significant alteration across cancer stages, indicating their roles in cancer progression and invasion. These lncRNAs were *SNHG12* and *SNHG20* in LIHC (Figure 3).

In addition to the lncRNAs, we found that several mRNA with worse prognosis in their corresponding networks showed significant altered expression across tumor stages, indicating their roles in cancer progression and invasion. These genes

were as follows: *LIMK1* ($p = 0.000934$), *PLXNA1* ($p = 0.00176$), *E2F8* ($p = 0.0155$), *EFNA3* ($p = 3.74e-07$), *SLC7A11* ($p = 0.0497$), *TMEM164* ($p = 0.00469$), *H2AFZ* ($p = 5.27e-06$), *BUB1* ($p = 5.82e-06$), *CBX2* ($p = 0.00033$), *EXO1* ($p = 4.88e-06$), *FANCE* ($p = 0.000301$), *LMNB2* ($p = 0.000311$), *SOX12* ($p = 0.00467$), *JPT1* ($p = 4.08e-07$), *MAFG* ($p = 0.00305$), *STK39* ($p = 0.0428$), *ANKRD13B* ($p = 0.000412$) (Supplementary Figure S3).

3.5 Providing Evidence That Highlights the Need for Functional Study on the Role of *KDELC1* in Cancers

The role of *RIPOR2*, *SMOC1*, *B3GNT5*, *ANKRD13B*, and *TMEM164* has not been clearly demonstrated in LIHC. Moreover, the role of *ZNF607* and mainly *KDELC1* has not been clearly delineated in any cancer [*in-silico* analysis by Zhou et al. (2021)]. However, our findings in this study highlight the need for functional study of *KDELC1* in cancers as follows: One—Our data showed its over-expression with worse prognosis (reduced OS) in HNSC. Two—It showed overexpression in several tumors such COAD, kidney renal clear cell carcinoma (KIRC), and kidney renal papillary cell carcinoma (KIRP) and lower expression in Kidney Chromophobe (KICH) (Table 4). Three—It is mainly localized in the nucleus (confidence score: 5) and endoplasmic reticulum (confidence score: 5), followed by cytosol (confidence score: 4) (<https://www.genecards.org> and <https://www.proteinatlas.org>). Four—It specifically targets extracellular EGF repeats of important proteins such as NOTCH1 and NOTCH3 (Takeuchi et al., 2018), indicating its possible roles in the Notch signaling pathway. Five—Using STRING database, we found that it had interactions with several proteins involved in different cancers such as *TXNDC9* (Feng et al., 2020b), *GXYLT2* (Cui et al., 2019), *POFUT1* (Chabanais et al., 2018), *XXYLT1* (Zeng et al., 2021), *FLNA* (Guo et al., 2018), and *ZC3H12C* (Suk et al., 2018) (Figure 4). These evidences strengthen the possible important role of *KDELC1* in various malignancies.



3.6 Functional Enrichment Analysis

In our study, firstly, we extracted the 10 top GO terms for BP, CC, and MF in the TCGA-GI cancers. Then, we compared these GO terms between GI cancers (Supplementary Figure S4 and

Supplementary Table S4). Our study revealed some common and specific GO terms for these cancers given in Supplementary Tables S5,6, respectively. These terms were previously shown in COAD cancer (Poursheikhani et al., 2020);

TABLE 3 | mRNAs as prognostic biomarkers identified in ceRNA networks of TCGA-GI cancers.

TCGA cancer type	Gene	Worse prognosis	Good prognosis
COAD	<i>HOMER1</i> <i>NRG1</i>		Over-expression
HNSC	<i>KDEL1C1</i> , <i>GNA12</i> , <i>ITGA5</i> , and <i>CDCA4</i> <i>MEIS1</i>	Low expression Over-expression Low expression	
LIHC	<i>RUNX3</i> , <i>CD69</i> , <i>RIPOR2</i> (<i>FAM65B</i>), <i>GADD45A</i> , <i>SYBU</i> , <i>ADAMTSL3</i> , <i>AUTS2</i> , <i>SMOC1</i> <i>SCML2</i> , <i>LIMK1</i> , <i>FANCE</i> , <i>PLXNA1</i> , <i>E2F8</i> , <i>EFNA3</i> , <i>SLC7A11</i> , <i>ENAH</i> , <i>TMEM164</i> , <i>H2AFZ</i> , <i>BUB1</i> , <i>CBX2</i> , <i>EXO1</i> , <i>LPL</i> , <i>B3GNT5</i> , <i>LMNB2</i> , <i>SOX12</i> , <i>JPT1</i> , <i>MAFG</i> , <i>ZNF607</i> , <i>STK39</i> , and <i>ANKRD13B</i> <i>CDC7</i> and <i>PFKFB3</i> <i>DNM3</i> and <i>COL15A1</i>	Low expression Over-expression	Low expression Over-expression
READ	<i>ATP8B2</i> , <i>NTN1</i> , <i>NOVA1</i> , and <i>PCDH7</i>		Low expression
STAD	<i>MAP7</i>		Over-expression

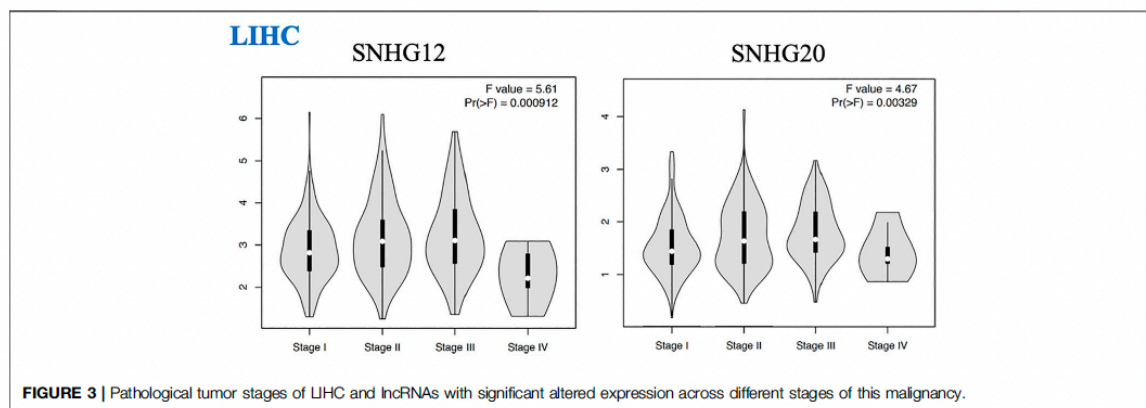


FIGURE 3 | Pathological tumor stages of LIHC and lncRNAs with significant altered expression across different stages of this malignancy.

TABLE 4 | *KDEL1C1* over-expression in TCGA cancers.

Cancer type	logFC	Average expression	p-value	Fdr
HNSC	1.4	2.62	7.05E-19	1.25E-17
COAD	1.1	2.38	1.79E-16	8.36E-16
KIRC	1.1	3.72	3.40E-24	1.61E-23
KIRP	1.95	4.35	5.58E-14	5.14E-13
KICH	-1.6	1.5	1.27E-16	2.18E-15

however, in our study, we looked for common and specific terms in GI cancers. For MF, our study revealed extracellular matrix structural constituent (GO:0005201) as the top shared term among all GI cancers (Supplementary Table S5). In relation to the CC, we identified collagen-containing extracellular matrix (GO:0062023), which was shared between all GI cancers (Supplementary Table S5). For BP, the most significantly shared GO terms in these cancers were extracellular matrix organization (GO:0030198), which was shared among all studied malignancies except LIHC. In addition, regulation of chromosome segregation (GO:0051983) and mitotic nuclear division (GO:0140014) were shared among all investigated cancers excepting READ (Supplementary Table S5). Regarding KEGG, we found Cell

cycle (hsa04110) as the top shared one among all GI cancers (Supplementary Table S5). However, we found that some BP, CC, and MF terms were exclusively specific for each cancer (Supplementary Table S6).

4 DISCUSSION

Our study revealed differential expression of mRNAs, lncRNAs, and miRNAs in ceRNA networks of GI cancers. We highlighted the important function of several lncRNAs and mRNAs specific for each cancer or commonly shared between GI cancers, which might be used for diagnostic and therapeutic purposes.

Our study showed that *MAGI2-AS3* lncRNAs were common between most of these cancers, representing possible involvement of the same pathways in different GI cancers. In ceRNA networks of each cancer, we also found that differential expression of some lncRNAs was specific to each cancer.

We found several lncRNAs as prognostic biomarkers including *LINC00958* and *LINC00707* in HNSC; *SNHG20* and *SNHG12* in LIHC (correlated with reduced OS with worse prognosis); *SNHG20* in READ; and *PVT1* and *MAGI2-AS3* in STAD (correlated with good prognosis). We found that over-expression of *SNHG12* and *SNHG20* in LIHC (Figure 3) had significant alteration across cancer

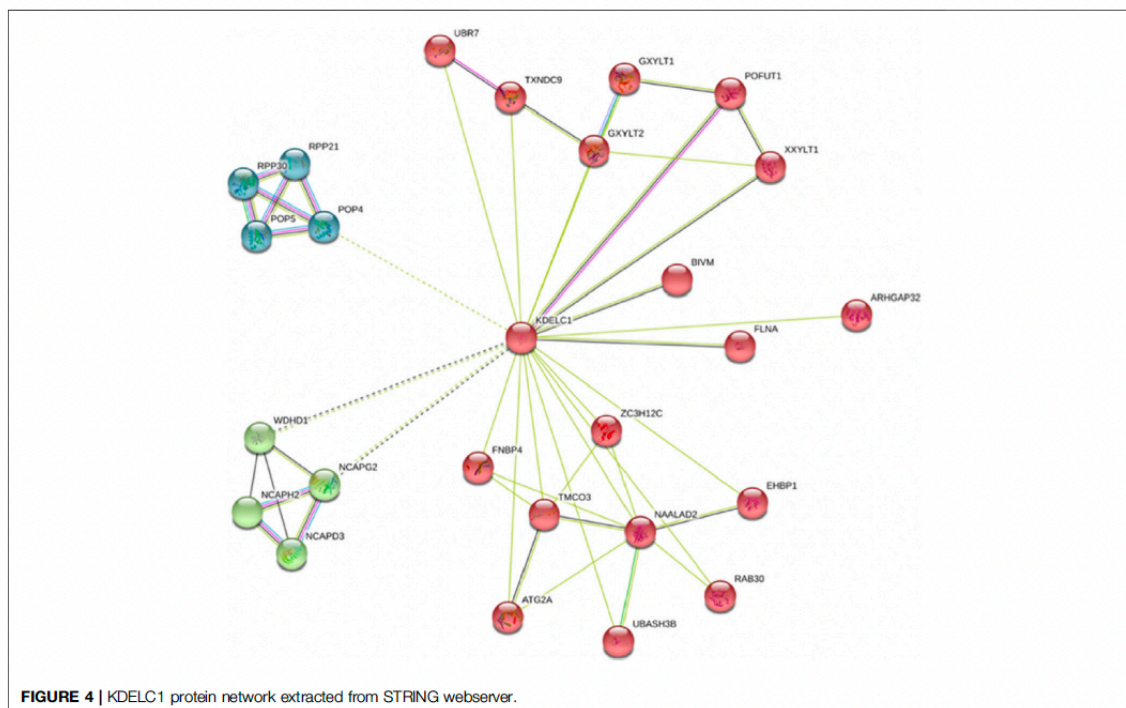


FIGURE 4 | KDEL1 protein network extracted from STRING webserver.

stages, representing their roles in cancer progression and invasion. These lncRNAs identified in our study were previously reported in their corresponding cancers (Yang et al., 2015; Zhang et al., 2016; Li et al., 2019; Tamang et al., 2019; Feng et al., 2020a; Niu et al., 2020a; Martinez-Barriocanal et al., 2020; Shen et al., 2020; Zeng et al., 2020; Zhang et al., 2020b; Zheng et al., 2020) but a comprehensive investigation of distinct or shared lncRNAs in these cancers has not yet been conducted.

Notably, we found a number of mRNAs with prognostic values (reduced OS with worse prognosis), which were not previously reported or supported by functional studies in their corresponding human tumors as follows: over-expression of *KDEL1* in HNSC; over-expression of *TMEM164*, *B3GNT5*, *ZNF607*, and *ANKRD13B*; and under-expression of *RIPOR2* and *SMOC1* in LIHC. The role of *ZNF607* and mainly *KDEL1* has not been clearly elucidated in cancers. Among these mRNAs, *TMEM164*, *ZNF607*, and *ANKRD13B* showed significant altered expression across tumor stages, indicating their role in cancer progression and invasion. Therefore, functional studies focused on these prognostic biomarkers in their corresponding cancers are suggested. Moreover, among GI cancers, four mRNAs were more common in most of them, which included *MEIS1* and *PPP1R3C* among COAD, READ, STAD, and HNSC; and *ADAMTSL3* and *MYLK* among COAD, READ, STAD, and LIHC. Furthermore, we found several important shared mRNAs between ceRNA networks of COAD and READ (Table 2, CRC, cancer cluster 3:CC-3).

According to the findings of the current study, we strongly proposed a novel proto-oncogene, *KDEL1* (*POGLUT2*), in

cancers since it is mainly localized in nucleus and endoplasmic reticulum, it specifically targets extracellular EGF repeats of important proteins involved in Notch signaling pathways, it has interactions with several main proteins involved in cancers, and it is upregulated in several TCGA cancers. Moreover, our study revealed several mRNAs associated with worse prognosis in cancer and showed their significant altered expression across tumor stages.

Our study found that in ceRNA networks of LIHC, combination of prognostic values of *SNHG20* lncRNA and *BUB1* mRNA, and also *SNHG12* lncRNA and *TMEM164* mRNA (associated with reduced OS with worse prognosis), could be used as collective prognostic values for this cancer.

In this study, we found extracellular matrix structural constituent as the top shared MF term, collagen-containing extracellular matrix as the top shared CC, and Cell cycle (*hsa04110*) as the top shared KEGG among all these GI cancers. However, extracellular matrix organization, except for LIHC, and regulation of chromosome segregation and mitotic nuclear division, both except for READ, were found to be the most significantly shared BP terms among the GI cancers.

In conclusion, we found several common and distinct prognostic mRNAs and lncRNAs in GI cancers. The most important finding was the involvement of *MAGL2-AS3* lncRNA in almost all GI cancers. Another interesting result was the identification of the most common shared mRNAs in these cancers. Moreover, our study revealed a number of specific lncRNAs and mRNAs in GI cancers. Furthermore, our study reported several genes, most notably *KDEL1*, that can have proto-oncogenic roles in

cancers. In addition to the abovementioned findings, we reported the top shared KEGG, GO-MF, GO-BP, and GO-CC among GI cancers. Overall, our study showed that bioinformatic studies regarding mRNA-miRNA-lncRNA networks of TCGA data can help identify prognostic biomarkers. The identified common shared and distinct mRNAs, lncRNAs, miRNAs, KEGG, and GO-terms among GI cancers can be used by other researchers for further functional experiments and also for diagnostic and prognostic approaches.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study include TCGA, GEPIA2, STRING, Genecards, and Human Protein Atlas. The raw data can be found here: <https://tcga-data.nci.nih.gov/tcga/>, <http://gepia2.cancer-pku.cn/#index>, <https://string-db.org>, <https://www.genecards.org>, and <https://www.proteinatlas.org>.

AUTHOR CONTRIBUTIONS

HD conceptualized and designed the study, performed bioinformatic analysis, wrote, reviewed and edited the manuscript. AA performed bioinformatic analysis, reviewed, and edited the manuscript. PH reviewed and edited the manuscript. AN reviewed and edited the manuscript. AH reviewed and edited the manuscript. YM reviewed and edited the manuscript. HH-A conceptualized the study, performed

systematic search for genes in article databases, reviewed and edited the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.791919/full#supplementary-material>

Supplementary Figure S1 | ceRNA networks of LIHC, HNSC, READ, and ESCA.

Supplementary Figure S2 | OS analysis of protein-coding genes in patients with TCGA-GI cancers (HNSC, LIHC, READ, STAD, COAD).

Supplementary Figure S3 | mRNAs with significant altered expression across tumor stages in ceRNA networks of GI cancers.

Supplementary Figure S4 | GSEA of GI cancers (COAD, ESCA, HNSC, LIHC, READ, STAD).

Supplementary Table S1 | Differentially expressed mRNAs, lncRNA, and miRNAs in the ceRNA networks of GI cancers.

Supplementary Table S2 | Common and distinct lncRNAs and specific mRNAs in ceRNA networks of GI cancers.

Supplementary Table S3 | Protein-coding genes involved in MAGI2-AS3 ceRNA network of GI cancers.

Supplementary Table S4 | The GO terms (BP, CC, and MF) and KEGG pathways involved in the TCGA-GI cancers.

Supplementary Table S5 | Common GO terms and KEGG pathways of the GI cancers.






Supplementary Table S6 | Specific and distinct GO terms and KEGG pathways of the GI cancers.

REFERENCES

- Arnold, M., Abnet, C. C., Neale, R. E., Vignat, J., Giovannucci, E. L., McGlynn, K. A., et al. (2020). Global Burden of 5 Major Types of Gastrointestinal Cancer. *Gastroenterology* 159, 335–349. e315. doi:10.1053/j.gastro.2020.02.068
- Chabanaï, J., Labrousse, F., Chauvel, A., Germot, A., and Maftah, A. (2018). POFUT1 as a Promising Novel Biomarker of Colorectal Cancer. *Cancers (Basel)* 10, 1. doi:10.3390/cancers10110411
- Cui, Q., Xing, J., Gu, Y., Nan, X., Ma, W., Chen, Y., et al. (2019). GXYLT2 Accelerates Cell Growth and Migration by Regulating the Notch Pathway in Human Cancer Cells. *Exp. Cell Res.* 376, 1–10. doi:10.1016/j.yexcr.2019.01.023
- Dastsooz, H., Cereda, M., Donna, D., and Oliviero, S. (2019). A Comprehensive Bioinformatics Analysis of UBE2C in Cancers. *Int. J. Mol. Sci.* 20, 1. doi:10.3390/ijms20092228
- Feng, L., Li, H., Li, F., Bei, S., and Zhang, X. (2020a). lncRNA KCNQ1OT1 Regulates microRNA-9-Lmx1a Expression and Inhibits Gastric Cancer Cell Progression. *Aging* 12, 707–717. doi:10.18632/aging.102651
- Feng, T., Zhao, R., Sun, F., Lu, Q., Wang, X., Hu, J., et al. (2020b). TXNDC9 Regulates Oxidative Stress-Induced Androgen Receptor Signaling to Promote Prostate Cancer Progression. *Oncogene* 39, 356–367. doi:10.1038/s41388-019-0991-3
- Guo, Y., Li, M., Bai, G., Li, X., Sun, Z., Yang, J., et al. (2018). Filamin A Inhibits Tumor Progression through Regulating BRCA1 Expression in Human Breast Cancer. *Oncol. Lett.* 16, 6261–6266. doi:10.3892/ol.2018.9473
- Iriondo-Dehond, A., Uranga, J. A., Del Castillo, M. D., and Abalo, R. (2020). Effects of Coffee and its Components on the Gastrointestinal Tract and the Brain-Gut Axis. *Nutrients* 13, 1. doi:10.3390/nu13010088
- Li, R., Qu, H., Wang, S., Wei, J., Zhang, L., Ma, R., et al. (2018). GDCRNATools: an R/Bioconductor Package for Integrative Analysis of lncRNA, miRNA and mRNA Data in GDC. *Bioinformatics* 34, 2515–2517. doi:10.1093/bioinformatics/bty124
- Li, S. J., Wang, L., Sun, Z. X., Sun, S. J., Gao, J., and Ma, R. L. (2019). lncRNA SNHG1 Promotes Liver Cancer Development through Inhibiting P53 Expression via Binding to DNMT1. *Eur. Rev. Med. Pharmacol. Sci.* 23, 2768–2776. doi:10.26355/eurrev_201904_17550
- Li, Z., Wang, F., Zhu, Y., Guo, T., and Lin, M. (2021). Long Noncoding RNAs Regulate the Radioresistance of Breast Cancer. *Anal. Cell Pathol (Amst)* 2021, 9005073. doi:10.1155/2021/9005073
- Martínez-Barriocanal, Á., Arango, D., and Dopeso, H. (2020). PVT1 Long Non-coding RNA in Gastrointestinal Cancer. *Front. Oncol.* 10, 38. doi:10.3389/fonc.2020.00038
- Niu, J., Song, X., and Zhang, X. (2020a). Regulation of lncRNA PVT1 on miR-125 in Metastasis of Gastric Cancer Cells. *Oncol. Lett.* 19, 1261–1266. doi:10.3892/ol.2019.11195
- Niu, Z.-S., Wang, W.-H., Dong, X.-N., and Tian, L.-M. -L. (2020b). Role of Long Noncoding RNA-Mediated Competing Endogenous RNA Regulatory Network in Hepatocellular Carcinoma. *Wjg* 26, 4240–4260. doi:10.3748/wjg.v26.i29.4240
- Poursheikhani, A., Abbaszadegan, M. R., Nokhandani, N., and Kerachian, M. A. (2020). Integration Analysis of Long Non-coding RNA (lncRNA) Role in Tumorigenesis of colon Adenocarcinoma. *BMC Med. Genomics* 13, 108. doi:10.1186/s12920-020-00757-2
- Sharma, K. L., Bhatia, V., Agarwal, P., and Kumar, A. (2018). Gastrointestinal Cancers: Molecular Genetics and Biomarkers. *Can. J. Gastroenterol. Hepatol.* 2018, 4513860. doi:10.1155/2018/4513860
- Shen, A., Ma, J., Hu, X., and Cui, X. (2020). High Expression of lncRNA-SNHG7 Is Associated with Poor Prognosis in Hepatocellular Carcinoma. *Oncol. Lett.* 19, 3959–3963. doi:10.3892/ol.2020.11490
- Suk, F. M., Chang, C. C., Lin, R. J., Lin, S. Y., Chen, Y. T., and Liang, Y. C. (2018). MCP1P3 as a Potential Metastasis Suppressor Gene in Human Colorectal Cancer. *Int. J. Mol. Sci.* 19, 1. doi:10.3390/ijms19051350

Hypothesis

Autophagy: The Potential Link between SARS-CoV-2 and Cancer

Parham Habibzadeh ¹, Hassan Dastsooz ^{2,3,4}, Mehdi Eshraghi ⁵, Marek J. Łos ^{6,*}, Daniel J. Klionsky ⁷ and Saeid Ghavami ^{5,8,9,4}

- ¹ Research Center for Health Sciences, Institute of Health, Shiraz University of Medical Sciences, Shiraz 71348-14336, Iran; Parham.Habibzadeh@yahoo.com
- ² Department of Life Sciences and Systems Biology, University of Turin, Via Accademia, Albertina, 13, 10123 Torino, Italy; Hassan.Dastsooz@unito.it
- ³ IIGM-Italian Institute for Genomic Medicine, c/o IRCCS, Candiolo, 10126 Torino, Italy
- ⁴ Candiolo Cancer Institute, FPO-IRCCS, 10060 Torino, Italy
- ⁵ Department of Human Anatomy and Cell Science, Rady Faculty of Health Sciences, Max Rady College of Medicine, University of Manitoba, Winnipeg, MB R3E 0J9, Canada; eshraghi.mehdi@gmail.com
- ⁶ Biotechnology Center, Silesian University of Technology, 44-100 Gliwice, Poland
- ⁷ Life Sciences Institute, University of Michigan, Ann Arbor, MI 48109, USA; klionsky@umich.edu
- ⁸ Research Institute of Oncology and Hematology, Cancer Care Manitoba, University of Manitoba, Winnipeg, MB R3E 0V9, Canada
- ⁹ Faculty of Medicine, Katowice School of Technology, ul. Rolna 43, 40-555 Katowice, Poland
- * Correspondence: Marek.Los@polsl.pl (M.J.Ł.); saeid.ghavami@umanitoba.ca (S.G.)



Citation: Habibzadeh, P.; Dastsooz, H.; Eshraghi, M.; Łos, M.J.; Klionsky, D.J.; Ghavami, S. Autophagy: The Potential Link between SARS-CoV-2 and Cancer. *Cancers* **2021**, *13*, 5721. <https://doi.org/10.3390/cancers13225721>

Academic Editors: Maria Dalamaga, Narjes Nasiri-Ansari and Nikolaos Spyrou

Received: 27 October 2021
Accepted: 14 November 2021
Published: 16 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Simple Summary: Coronavirus disease 2019 (COVID-19) has led to a global crisis. With the increasing number of individuals infected worldwide, the long-term consequences of this disease have become an active area of research. The constellation of symptoms COVID-19 survivors suffer from is commonly referred to as post-acute COVID-19 syndrome in the scientific literature. In this paper, we discuss the potential long-term complications of this infection resulting from the persistence of the viral particles in body tissues interacting with host cells' autophagy machinery in the context of the development of cancer, cancer progression and metastasis, as well as response to treatment. We also propose a structured framework for future studies to investigate the potential impact of COVID-19 infection on cancer.

Abstract: COVID-19 infection survivors suffer from a constellation of symptoms referred to as post-acute COVID-19 syndrome. However, in the wake of recent evidence highlighting the long-term persistence of SARS-CoV-2 antigens in tissues and emerging information regarding the interaction between SARS-CoV-2 proteins and various components of the host cell macroautophagy/autophagy machinery, the unforeseen long-term consequences of this infection, such as increased risk of malignancies, should be explored. Although SARS-CoV-2 is not considered an oncogenic virus, the possibility of increased risk of cancer among COVID-19 survivors cannot be ruled out. Herein, we provide an overview of the possible mechanisms leading to cancer development, particularly obesity-related cancers (e.g., colorectal cancer), resulting from defects in autophagy and the blockade of the autophagic flux, and also immune escape in COVID-19 survivors. We also highlight the potential long-term implications of COVID-19 infection in the prognosis of patients with cancer and their response to different cancer treatments. Finally, we consider future directions for further investigations on this matter.

Keywords: colorectal neoplasms; COVID-19; gastrointestinal neoplasms; immune checkpoint inhibitors; neoplasms; oncogenic viruses; oncolytic virotherapy; post-acute COVID-19 syndrome; reactive oxygen species; tumor escape

1. Introduction

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has led to unprecedented mortality and morbidity at a global scale. SARS-CoV-2 is a positive-stranded RNA virus belonging to the *Coronaviridae* family, members of which interact with different components of the cellular autophagy machinery [1,2]. A constellation of symptoms, such as fatigue, exhaustion, and shortness of breath, persisting long after the resolution of the acute phase of infection, has been reported among some survivors of this infection [3]. The persistence of symptoms beyond 12 weeks after the initial onset is called post-COVID-19 syndrome [3,4]. However, long-term complications of this infection could extend well beyond these symptoms.

2. Autophagy and Cancer

Autophagy plays a prominent role in maintaining cellular homeostasis through the removal of damaged organelles, abnormal proteins, and invading organisms. Defects in autophagy are associated with various pathological conditions, including cancer [5]. It can lead to accumulation of damaged mitochondria and alter cellular metabolism, leading to a high oxidative state [6]. Furthermore, impairments in autophagy flux can lead to ER stress and subsequent accumulation of chaperone proteins and an eventual rise in the unfolded protein burden [7,8]. This chronic injury to various cellular organelles and proteins and the accumulation of the genetic damage in cells with defective autophagy can lead to the development of pathophysiologies, as evidenced by the detection of loss-of-function mutations in different autophagy genes in various malignancies such as colorectal cancer [9]. However, despite this significant role of autophagy in the initiation of tumors, it is a double-edged sword in cancer, and its role is highly context dependent [10]. Cancer cells are dependent on the cellular autophagy machinery due to their rapid growth and biosynthetic demands [6,8]. Autophagy also plays a significant role in promoting metastasis in certain tumors, particularly in RAS-driven cancers [11]. Conversely, autophagy can also be leveraged to enhance the response to various cancer treatments [12] (Figure 1).

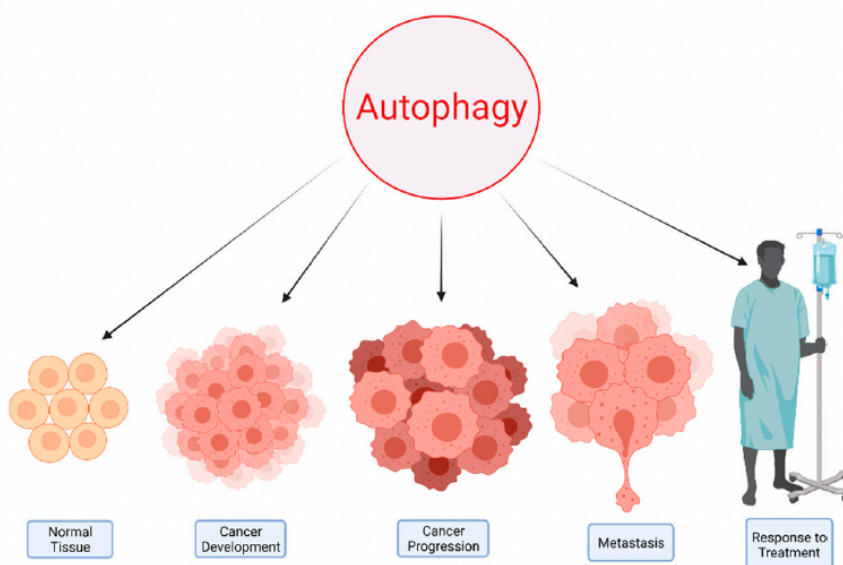


Figure 1. Autophagy has a principle role in both the health and disease states, such as malignancies. This process plays a complex role in various aspects of cancer.

3. Potential Clues and Mechanisms Supporting the Role of SARS-CoV-2 in Oncogenesis

Several oncogenic viruses exert their carcinogenesis through altering autophagy [13]. Although the oncogenic potential of SARS-CoV-2 has not yet been investigated, two other positive-sense single-strand RNA viruses, namely, hepatitis C virus/HCV and human T-cell lymphotropic virus type 1/HTLV-1, exploit the cellular autophagy machinery in order to cause liver cancer and adult T-cell leukemia/lymphoma/ATLL, respectively [13].

Both ACE2 (the major SARS-CoV-2 receptor) and TMPRSS2 (a transmembrane serine protease necessary for viral cell entry) display a very high level of expression in the human gastrointestinal tract [14]. This virus can infect and actively replicate in human enterocytes [14]. A recent study on the gastrointestinal biopsies of 14 individuals performed four months after their COVID-19 diagnosis reported persistence of the viral nucleic acids and antigens in 50% of the cases [15]. Persistent shedding of SARS-CoV-2 viral particles for months after the resolution of the symptoms has also been reported in the nasopharyngeal specimens obtained from individuals with previous infection [16,17]. This continued presence of the viral antigens in different tissues could have serious long-term consequences for survivors of COVID-19 infection.

Interestingly, various SARS-CoV-2 proteins interact with different components of the cellular autophagy pathway (Figure 2). For example, viral ORF3a protein interacts with VPS39, a part of the homotypic fusion and protein sorting/HOPS complex, leading to the inhibition of the fusion of autophagosomes with lysosomes [18]. In addition, SARS-CoV-2 Nsp15 blocks the induction of autophagy, whereas the viral ORF7a protein decreases the acidity of lysosomes, which can interfere with autophagosome–lysosome fusion as well as cargo degradation [19]. Moreover, accumulation of SQSTM1 and an increase in the processed LC3B (LC3B-II) levels have been observed upon SARS-CoV-2 envelope (E) protein overexpression [19,20]. Altogether, these findings indicate that various SARS-CoV-2 antigens block autophagic flux in the infected cells. Conversely, a recent study showed that the protein encoded by viral ORF8 leads to major histocompatibility complex I (MHC-I) degradation in the affected cells by targeting these molecules for lysosomal degradation through the BECN1-dependent autophagy pathway [21].

The persistence of the SARS-CoV-2 antigens in the enterocytes could lead to long-term defects in the cellular autophagy machinery, as evidenced by recent studies. The blockage in the autophagic flux in these cells would lead to the accumulation of SQSTM1/p62 (an autophagic receptor protein), reactive oxygen species, organelle damage, and genetic alterations in response to different stresses, eventually leading to tumorigenesis [22,23]. An approximate 1.5-fold increase in SQSTM1 levels in the presence of SARS-CoV-2 ORF3a, ORF7a, or E proteins has been observed [19]. SQSTM1 plays a significant role in tumor transformation due to its important function as a signaling molecule interacting with many oncogenic pathways, including those involving NFE2L2/NRF2 and NFKB/NF- κ B [24]. Several studies also reported evidence of mitochondrial dysfunction, excessive production of reactive oxygen species, ER stress, and unfolded protein responses in the cells infected with SARS-CoV-2 [25–27]. As mentioned, these pathological events could be, in part, a result of disturbed autophagy flux in the cells. Furthermore, disrupted cell cycle regulation due to autophagy defects would lead to the uncontrolled proliferation of cells carrying defective genetic materials, paving the way for the development of cancer [28]. In addition, activation of compensatory mechanisms in colorectal cancer cells in specific contexts secondary to the inhibition of autophagy lead to tumor growth [29]. This could in turn lead to a higher risk of cancer development and more rapid proliferation of tumor cells among COVID-19 survivors.

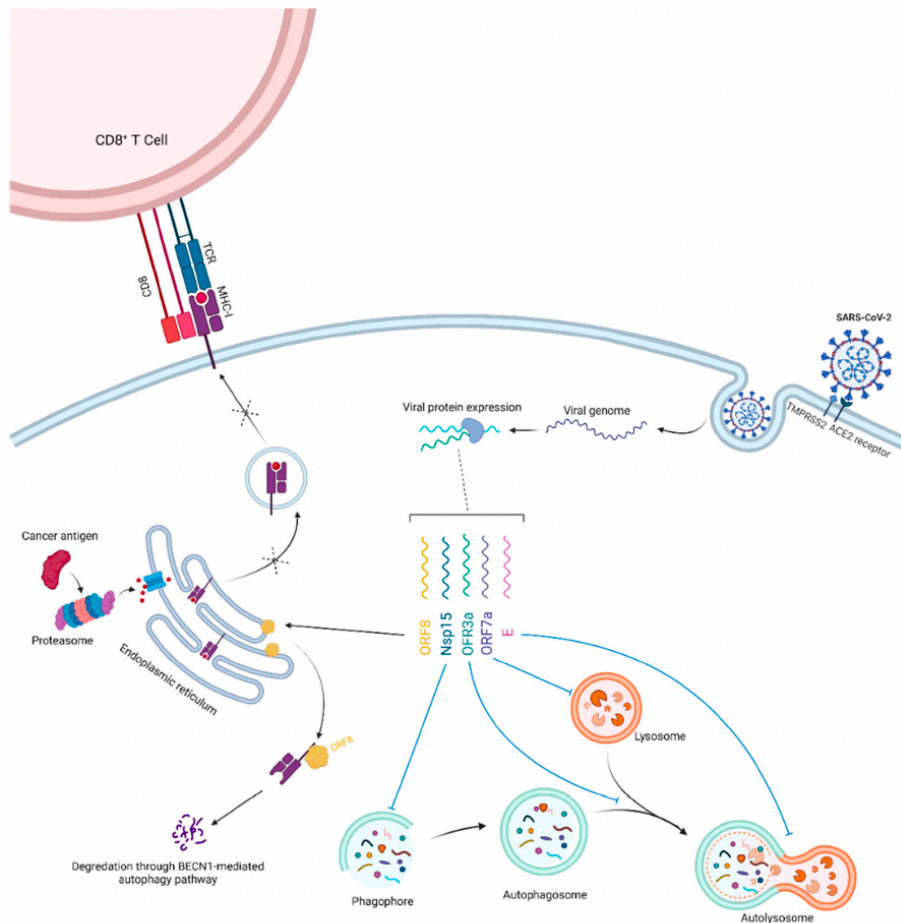


Figure 2. SARS-CoV-2 infects enterocytes through ACE2 receptors and TMPRSS2 expressed on their surface. The release of the viral genome and production of viral proteins (simplified in this schema) will lead to the interaction of various SARS-CoV-2 antigens with the host cell autophagy machinery. Overall, this will induce the blockage of the autophagic flux. The accumulation of reactive oxygen species, damaged cellular proteins and organelles, and acquired genetic defects due to various stressors can lead to the development of malignancies. Furthermore, infection can also interfere with the cellular MHC-I antigen-presentation pathway, blunting the ability of host cytotoxic CD8⁺ T cells to recognize potential oncogenic antigens.

Downregulation of MHC-I is one of the major mechanisms of the immune evasion by tumor cells through escaping detection by CD8⁺ T-cells and their cytotoxicity [30]. This adaptive immune escape mechanism is observed in many malignancies, including colorectal cancer, and is associated with a poor prognosis [31]. MHC-I degradation through the BECN1-dependent autophagy pathway induced by viral ORF8 could, therefore, provide a fertile soil for carcinogenesis by blunting the ability of the immune system to detect the cancer cells because their neo-antigens are no longer presented on the MHC-I molecules.

It should be noted that SARS-CoV-2 antigens have also been detected in various other human organs, such as the lungs, heart, kidneys, hepatobiliary system, and the lymphatic system [32]. Although most of these findings have been reported in post-mortem studies and the persistence of the viral antigens in these organs among COVID-19 survivors has

not so far been assessed due to technical and ethical considerations, it is likely that viral particles could linger in these tissues, continuing to interact with the host cell autophagy machinery and therefore inducing carcinogenesis in various organ systems. Chronic SARS-CoV-2 infection is reported in immunocompromised patients (e.g., due to anti-cancer therapy) and animal models [33–35]. These immunocompromised patients with chronic infection might have a higher risk of cancer, as evidenced in other malignancies with suspected infectious etiology [36].

Interestingly, a recent study has reported that SARS-CoV-2 RNA can integrate into the genome of cultured human cells through reverse transcription. The authors also claimed that they were able to detect viral-host chimeric transcripts in the patient-derived tissues, suggesting that these transcripts are a result of integration of DNA copies of viral sequences in the human genome [37]. However, a later study by another group did not find any evidence for such an event in human cells [38]. Clearly, further large-scale studies are required to investigate if the SARS-CoV-2 genome is able to integrate into the human genome.

A nation-wide population-based study conducted in Denmark investigating mortality rates among patients admitted to the hospitals for non-COVID-19 diseases during the pandemic from March 2019 to January 2021, and found a consistently higher mortality rate among patients with cancer compared with baseline pre-pandemic mortality rates among these patients [39]. Although this observation is likely due to a multitude of factors, it is possible that defects in autophagy and increased tumor immune evasion among patients with certain malignancies who had previously contracted this infection could have led to more a rapid progression of their cancer due to the processes described above. In addition, a 6.9-fold increase in tumor burden has been reported in patients who had been diagnosed with metastatic colorectal cancer after the first lockdown compared with those diagnosed prior to the lockdown [40]. Despite the significant role of the delays in the screening and diagnosis as a result of the COVID-19-related lockdown, the above-mentioned processes could have also contributed to this observation. Notably, patients with cancer remain a vulnerable population for COVID-19 [41].

4. Autophagy and Metabolism

Autophagy plays a fundamental role as a catabolic process recycling intracellular components into breakdown products that could be used in various cellular metabolic pathways [42]. This, in turn, enables cells to survive under metabolic stress conditions (e.g., nutrient deprivation, hypoxia, etc.) [43,44].

Alteration in cellular metabolism is one of the hallmarks of cancer [45]. Unlike normal cells, which mainly rely on mitochondrial oxidative phosphorylation for energy production, most cancer cells depend on aerobic glycolysis (the Warburg effect) [45,46]. Despite being an inefficient pathway for adenosine 5'-triphosphate (ATP) generation, it is thought that it confers an advantage to tumor cells in incorporating nutrients in tumor biomass and cellular proliferation [46]. Considering the diverse role of autophagy in feeding different metabolic pathways through degradation of various cellular substrates, it is not surprising that elevated basal autophagy is seen in many tumor types [8,47].

Furthermore, mitochondrial metabolism plays an important role in various tumor types by redox balance, ATP generation, and synthesis of intermediates required for macromolecule biosynthesis (e.g., nucleotides) [48]. The selective autophagic elimination of the mitochondria (i.e., mitophagy) is a very important cellular process regulating their number and also participating in mitochondria quality control [49]. In addition, mitophagy plays a vital role in the metabolic rewiring of cancer cells aimed at meeting their bioenergetic needs [50].

Considering the paramount role of autophagy in maintaining tissue homeostasis, dysregulation in this process has been linked to cancer development and progression in a context-specific manner [51]. The potential inhibition of the autophagy flux by SARS-CoV-2 antigens can deprive tumor cells of the building blocks essential for unconstrained tumor

proliferation and could, therefore, limit tumor growth. However, the accumulation of dysfunctional organelles, particularly dysfunctional mitochondria, as a result of impaired autophagy can pave the way for the development of cancer [52] (Figure 3).

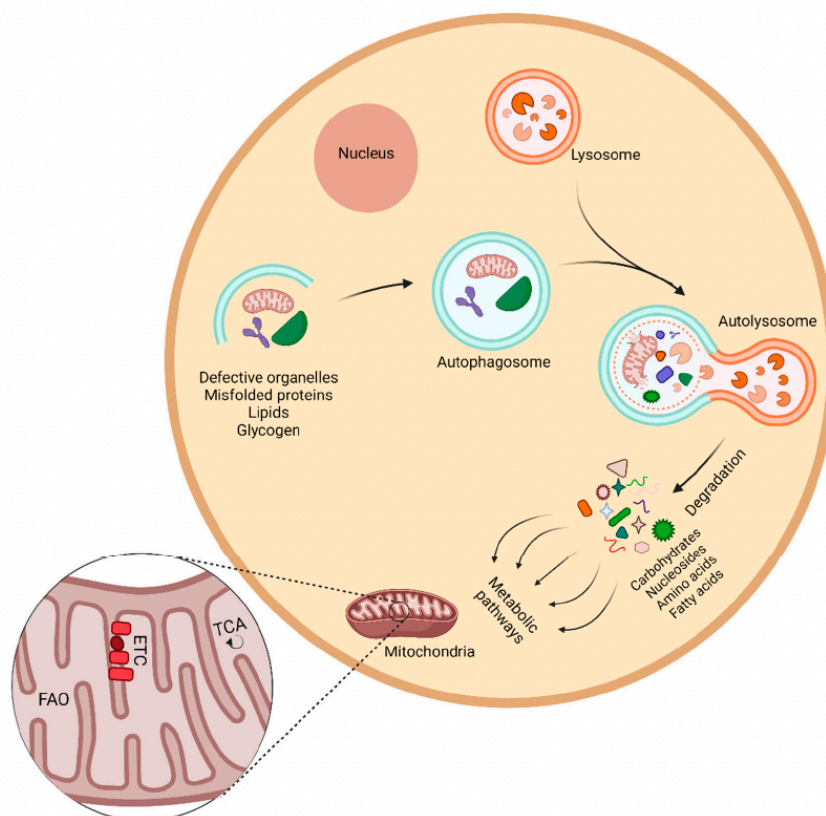


Figure 3. Autophagy and cell metabolism in cancer. Autophagy plays an important role in removing defective cellular organelles and macromolecules and recycling them for further use in different metabolic pathways. Mitochondria are the primary cellular organelles responsible for various cellular metabolic pathways including fatty acid oxidation (FAO), the tricarboxylic acid (TCA) cycle, and the electron transport chain (ETC). Impaired production of metabolic intermediates important for cellular proliferation can lead to a metabolic crisis and subsequently limit tumor growth. Conversely, excessive production of reactive oxygen species and the resultant oxidative damage can lead to tumorigenesis.

Obesity is correlated with elevated systemic oxidative stress [53]. The excessive nutrients supply, overwhelming the cellular Krebs cycle and mitochondrial respiratory chain, leads to mitochondrial dysfunction and increases the formation of reactive oxygen species (ROS) [54]. Elevation of intracellular ROS is known to lead to the upregulation of the cellular autophagic response, which subsequently removes defective mitochondria and therefore limits the generation of ROS [55–57]. The potential inhibition of autophagic flux as a result of the persistence of SARS-CoV-2 antigens in different tissues could blunt this protective mechanism and could particularly be more important in the pathogenesis of obesity-related cancers. ROS levels are higher in colorectal cancer cells compared with normal non-cancerous tissues [58]. Furthermore, ROS play a critical role in mediating tumorigenesis and colorectal cancer initiation driven by RAC1 [59]. The effects of ROS extend beyond the initiation of obesity-related cancers; they are anti-apoptotic factors promoting the survival of pancreatic cancer cells [60].

5. Implications for Cancer Treatment

The long-term presence of SARS-CoV-2 antigens in cancer tissues could also have wide-ranging implications for cancer therapy. MHC-I loss or downregulation is a common mechanism for the development of resistance to PD-1/PD-L1 inhibitors among patients with melanoma [61,62]. The same phenomenon could potentially render cancer cells arising in the presence of SARS-CoV-2 antigens resistant to anti-PD-1 monotherapy as a result of MHC-I degradation through BECN1-dependent autophagy. This outcome could necessitate the use of PD-1 and CTLA4 blockade combination therapy in such cases, as this treatment strategy does not require MHC-I expression for exerting its therapeutic effects [63].

Lysosomal sequestration of weak base hydrophobic chemotherapeutic agents decreasing their accessibility to their target sites and the resultant decrease in their cytotoxic effects is a significant challenge that culminates in treatment failure and subsequent increase in mortality due to cancer [64]. Lysosomal exocytosis facilitating the clearance of chemotherapeutics accumulated in this organelle is another important component of lysosome-mediated chemotherapy resistance [65]. A recent study has shown that SARS-CoV-2 ORF3a not only inhibits the autophagy flux but also promotes lysosomal exocytosis [66]. This could confer drug resistance to the cancer cells arising in the setting of persistence SARS-CoV-2 ORF3a presence.

Despite the preliminary clues pointing towards the role of the virus in promoting tumor progression described here, rare instances of tumor burden reduction in three patients with colorectal cancer and disease remission in a patient with Hodgkin lymphoma following COVID-19 infection have been reported in the literature [67,68]. Although the over-exuberant immune response instigated by SARS-CoV-2 attacking the tumor cells could have led to this phenomenon [69], the complex role of autophagy in cancer (acting as a double-edged sword) should not be underestimated in interpreting these exceptional cases. These serendipitous observations may partly be due to the blockage of autophagy flux, leading to the deprivation of cancer cells of the essential biosynthetic materials generated by the cellular autophagy machinery required for tumor growth [6,8]. This in turn highlights the therapeutic potential of SARS-CoV-2 in oncolytic virotherapy via blockage of autophagic flux in specific tumor types at certain stages of their natural course. Theoretically, this potential therapeutic strategy could particularly be effective against tumor cells with high autophagy activity.

6. Future Directions

Notwithstanding significant efforts to unravel the role of autophagy in COVID-19 infection, many questions, particularly those surrounding the role of autophagy in the long-term complications of COVID-19, remain unanswered [70,71]. A multidisciplinary approach, involving both clinical and basic science researchers, aimed towards studying the unforeseen long-term consequences of COVID-19 infection in cancer development, tumor progression, metastasis, and response to various cancer therapeutics could eventually piece together different parts of this puzzle, providing a better understanding of the complex interaction between SARS-CoV-2 and cancer.

The potential long-term oncogenic effects of SARS-CoV-2 antigens that inhibit the autophagic flux (e.g., NSP15, ORF3a, etc.) could be investigated through *in vitro* studies assessing changes in tumor formation following the long-term presence of these antigens in various human cell lines, such as human *KRAS* knockout cells. In addition, studies on patient-derived cancer cell lines could provide valuable insight into the effect of viral antigens and the subsequent blockade of the autophagic flux and MHC-I downregulation on tumor cell response to various cancer treatments and its effect on metastatic potential. Furthermore, although due to the short lifespan of most animal models they may not be suitable for investigating the long-term effects of viral antigens on cancer development, patient-derived mouse xenograft models of different human malignancies that have arisen in the presence of SARS-CoV-2 antigens can provide an ideal framework for exploring

tumor growth, metastasis, and the response to different cancer treatments. Comprehensive molecular phenotyping of various human tumors arising in the setting of long-term presence of SARS-CoV-2 antigens and comparing them with malignancies developed in other settings could shed light on the potential biological processes specifically affected (Figure 4).

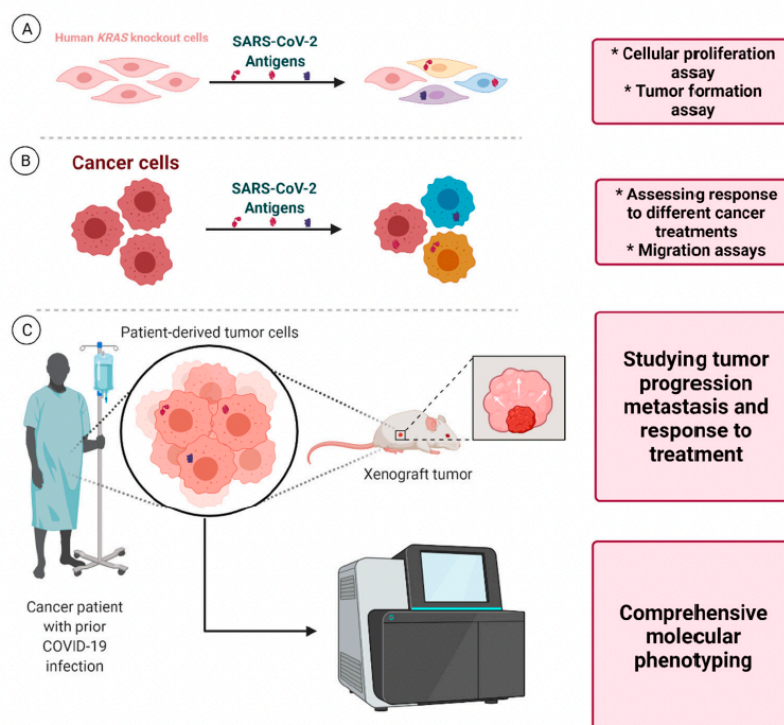


Figure 4. Proposed framework for laboratory-based investigations. (A) Investigations that could unravel the potential effects of various SARS-CoV-2 antigens on cellular proliferation and tumor formation in human KRAS knockout cells. (B) Using patient-derived cancer cells, the effect of these antigens and the subsequent autophagy blockade and MHC-I downregulation on response to different cancer treatments and metastasis (migration assays) could be studied. (C) Furthermore, studying tumor specimens arising in the presence of SARS-CoV-2 antigens using xenograft mouse models and also advanced molecular phenotyping techniques could unravel tumor biological behavior (growth rate, metastasis, etc.), its molecular characteristics, and the underlying factors affecting its response to various cancer therapeutics.

Clinical and epidemiological investigations can also play a principal role in expanding our understanding of this matter. Cohort studies with sufficient follow-up can assess the long-term incidence of various malignancies in different populations and would determine whether prior COVID-19 infection is a risk factor for the development of different cancers adjusting for well-known cancer risk factors. In addition, studying specific subpopulations, such as those with severe infection (i.e., ICU admission), and also patients with primary and secondary immunodeficiency could offer profound insight into the most susceptible populations. Considering the potential far-reaching implications of the infection for patients with active cancer, specific studies focused on this vulnerable population assessing the long-term impact of the infection on the clinical course of their malignancy should be carried out. Prospective studies investigating if prior COVID-19 infection affects the rate of

cancer progression, mortality rate, and response to treatment among patients with different malignancies, adjusting for the traditional contributing factors, could provide insight into the possible interaction between SARS-CoV-2 antigens and neoplastic cells (Figure 5).

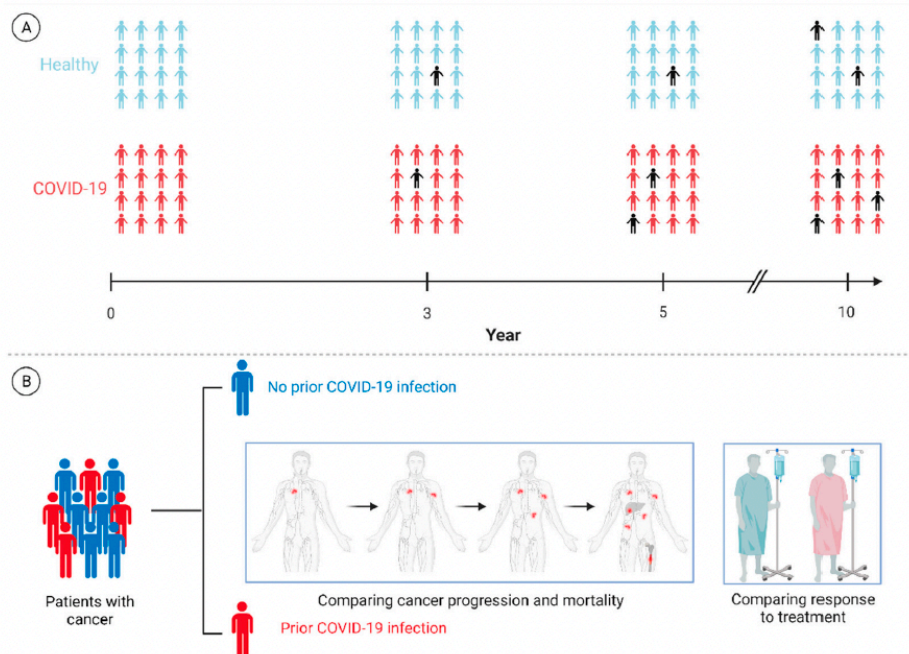


Figure 5. Proposed framework for clinical and population-based investigations. (A) Studies that can investigate whether prior COVID-19 infection is a risk factor for development of malignancies in subsequent years. (B) Other investigations could also shed light on the potential long-term effects of SARS-CoV-2 antigens present in cancer tissues on cancer progression and mortality rate and the response to different cancer therapeutics. The back icon in (A) represents a patient with a new diagnosis of cancer. Please note that the numbers of individuals with cancer in both groups have been arbitrarily chosen and are not based on real-life data. Future investigations will hopefully examine the hypothesis described here.

7. Conclusions


Despite the short amount of time since SARS-CoV-2 was first reported in 2019, scientists around the globe have managed to unravel various aspects of this infectious disease. However, we are still far from a solid understanding of this emerging infection and many important questions, particularly those regarding the long-term complications of this disease, remain unanswered.

In this work, building on the previous investigations demonstrating long-term persistence of the SARS-CoV-2 nucleic acids and antigens in human tissues and also other research studies showing the interaction of the viral particles with the host autophagy machinery, we hypothesize that SARS-CoV-2 could potentially be an oncogenic virus by blocking the autophagic flux, and also leading to immune escape by downregulation of MHC-I. We also propose that the resultant dysregulation in cellular autophagy could affect the response to treatment in cancer cells. Further laboratory-based, clinical, and population-based studies are required to explore this matter.

Author Contributions: Conceptualization, P.H., S.G.; resources, S.G. and M.J.L.; writing—original draft preparation, P.H. and H.D.; writing—review and editing, P.H., M.E., D.J.K., M.J.L. and S.G.; visualization, P.H., S.G.; supervision, S.G. and M.J.L. All authors have read and agreed to the published version of the manuscript.

Clinical and molecular characterization of a patient with mitochondrial Neurogastrointestinal Encephalomyopathy



Parham Habibzadeh^{1,2}, Mohammad Silawi¹, Hassan Dastsooz^{1,3}, Shima Bahramjahan¹, Shahrokh Ezzatzadegan Jahromi⁴, Vahid Reza Ostovan⁵, Majid Yavarian¹, Mohammad Mofatteh⁶ and Mohammad Ali Faghihi^{1,7*} 

Abstract

Background: Mitochondrial neurogastrointestinal encephalomyopathy (MNGIE) is a rare autosomal recessive disorder caused by mutations in *TYMP* gene, encoding nuclear thymidine phosphorylase (TP). MNGIE mainly presents with gastrointestinal symptoms and is mostly misdiagnosed in many patients as malabsorption syndrome, inflammatory bowel disease, anorexia nervosa, and intestinal pseudo-obstruction. Up to date, more than 80 pathogenic and likely pathogenic mutations associated with the disease have been reported in patients from a wide range of ethnicities. The objective of this study was to investigate the underlying genetic abnormalities in a 25-year-old woman affected with MNGIE.

Case presentation: The patient was a 25-year-old female referred to our center with the chief complaint of severe abdominal pain and diarrhea for 2 years that had worsened from 2 months prior to admission. The clinical and para-clinical findings were in favor of mitochondrial neurogastrointestinal encephalomyopathy syndrome. Subsequent genetic studies revealed a novel, private, homozygous nonsense mutation in *TYMP* gene (c. 1013 C > A, p.S338X). Sanger sequencing confirmed the new mutation in the proband. Multiple sequence alignment showed high conservation of amino acids of this protein across different species.

Conclusion: The detected new nonsense mutation in the *TYMP* gene would be very important for genetic counseling and subsequent early diagnosis and initiation of proper therapy. This novel pathogenic variant would help us establish future genotype-phenotype correlations and identify different pathways related to this disorder.

Keywords: Mitochondrial diseases, Mitochondrial neurogastrointestinal encephalomyopathy syndrome, *TYMP*, Codon, nonsense

* Correspondence: MFaghihi@med.miami.edu

¹Persian BayanGene Research and Training Center, Shiraz University of Medical Sciences, Shiraz, Iran

⁷Department of Psychiatry and Behavioral Sciences, University of Miami Miller School of Medicine, Miami, USA

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Mitochondrial neurogastrointestinal encephalomyopathy (MNGIE – OMIM# 603041) is a rare multisystem autosomal recessive disorder caused by homozygous or compound heterozygous mutations in the nuclear-encoded thymidine phosphorylase gene (*TYMP*; 131, 222) on chromosome 22q13, the first gene whose role was defined at molecular level in the defects of inter-genomic communication [1]. Impairment in this enzyme with resultant decreased enzyme activity leads to accumulation of the enzyme's substrates, thymidine and deoxyuridine, which in turn leads to an imbalance in the intra-mitochondrial nucleotide pool and multiple deletions, point-specific mutations and depletions in mitochondrial DNA (mtDNA) [2, 3]. Earlier studies have suggested that the accumulation of these nucleosides is the main culprit for the development of the molecular and phenotypic aberrations reported in this disorder [4–6].

First described in 1976, MNGIE usually presents before the age of 30 years with a mean age of onset of 18.5 years. Nevertheless, there have been a few patients where onset was as early as 5 months of age and as late as the fifth decade [7–9]. The disease has a progressive clinical course leading to death at the mean age of 37 [9–12]. The clinical manifestations of MNGIE are severe gastrointestinal dysmotility, cachexia, extraocular muscle weakness with resultant ptosis and ophthalmoplegia, sensorimotor neuropathy, and leukoencephalopathy [13]. More than 120 patients with diagnostic features consistent with MNGIE have so far been reported in the literature, with more than 80 pathogenic and likely pathogenic mutations (<https://www.ensembl.org>) associated with the disease identified in patients from a wide range of ethnicities [14, 15].

Herein, we report on a patient with MNGIE with a novel homozygous mutation in *TYMP* gene, along with the clinical, laboratory and imaging findings.

Case presentation

Clinical presentation

A 25-year-old female was referred to our center with the chief complaint of severe abdominal pain and diarrhea for 2 years that had worsened from 2 months prior to admission. She had significant weight loss during this period; weighing 36.5 kg with a height of 160 cm, her body mass index (BMI) was 14.3 kg/m² at the time of admission. Her past medical and surgical history was only significant for one undocumented episode of seizure at the age of three and appendectomy 3 years before. On interview, she denied fear of weight gain, laxative abuse, and self-induced vomiting. Her parents were consanguineous. There was no history of sibling loss or any similar symptoms in other family members.

Clinical examination revealed a cachectic lady with external ophthalmoplegia, ptosis, right lower quadrant scar of the McBurney (oblique) incision for appendectomy, decreased muscle power in the upper (4/5 MRC muscle scale) and lower (3/5 MRC muscle scale) extremities, and absent deep tendon reflexes. Abdominopelvic sonography revealed mild free fluid in the abdominal cavity and increased thickness in the bowel wall (4.8 mm) and a 16 × 11-mm cortical cyst in the upper pole of the left kidney with thin septation. A diagnostic esophago-gastro-duodenoscopy showed diffuse severe erythema and congestion in the body, fundus, and antrum of the stomach with moderate chronic gastritis in pathologic examination and deformity of the duodenal bulb with decreased folds in D2 part of the duodenum. On colonoscopy, the ileocecal valve was stenotic; biopsy revealed submucosal fibrosis with lymphoid proliferation and focal ulceration. Echocardiography was normal except for mild pericardial effusion.

Her complete blood count indicated the presence of microcytic hypochromic anemia. Biochemistry revealed low total protein level (4.1 g/dl; reference range: 6.6–8.8 g/dl) and albumin (2.2 g/dl; reference range: 3.5–5.2 g/dl). Serum lactate level (35.3 mg/dl; reference range: 4.5–19.8 mg/dl) was elevated. Fecal occult blood test was positive with moderately elevated fecal calprotectin level (58.5 µg/g; reference range: < 15 µg/g) suggestive for the inflammatory process. Cerebrospinal fluid analysis revealed marked elevation of protein level (122 mg/dl; reference range: 15–45 mg/dl). Serology tests for HIV, HBV, and HCV were negative. Serum anti-tissue trans-glutaminase antibodies, anti-phospholipid antibodies, anti-cardiolipin antibodies, lupus anti-coagulants, β₂ microglobulin, anti-nuclear antibody (ANA), anti-double stranded DNA (anti-dsDNA), C-ANCA, and P-ANCA were also within the normal limits. Serum levels of complement factors were also found to be altered: C₃ level (65 mg/dl; reference range: 90–180 mg/dl) was decreased; C₄ level (11.4 mg/dl; reference range: 10–40 mg/dl) was within the lower limit of the normal range. Blood TP activity was not measured for lack of laboratory resources.

The electrodiagnostic evaluation showed a neurogenic pattern on needle electromyography (EMG), conduction block in sensory nerves, and decreased compound muscle action potential (CMAP) in motor nerves with decreased conduction velocity and prolonged F-wave latency. Brain MRI with contrast showed leukoencephalopathy with diffusely increased T₂ signal intensity in both cerebral hemispheres white matter. Hypersignal intensity in splenium of corpus callosum was also observed (Fig. 1).

Molecular analysis

After obtaining informed consent, a blood sample was obtained from the patient. The whole blood sample was

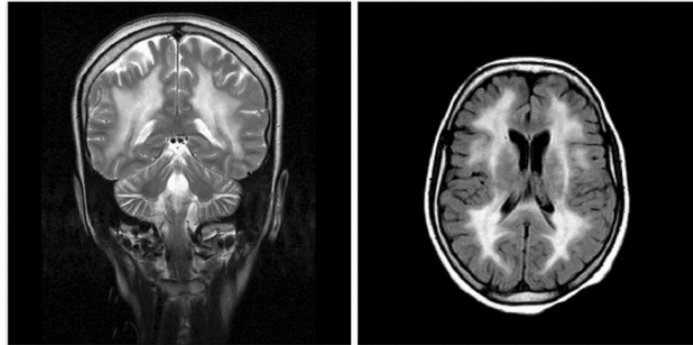


Fig. 1 Brain MRI revealing diffuse white matter signal changes

collected in EDTA tube. Genomic DNA was then extracted from peripheral leukocytes using the QIAamp DNA Blood Mini Kit (Qiagen, Germany).

Whole Exome Sequencing (WES) was performed on Illumina NextSeq500 instrument to a sequence close to 100 million reads. WES result was then analyzed using different bioinformatics tools and databases such as BWA aligner, GATK and ANNOVAR [16–18]. WES uncovered a novel, private, homozygous stopgain mutation in *TYMP* gene (NM_001113756: exon7: c. 1013C > A: p.S338X, Chr: 50526392).

Subsequently in order to confirm this novel mutation using Sanger sequencing, the region of interest was amplified using PCR on the DNA of the proband using following primers: F–*TYMP*-E7: 5'-ACTTAAGGGACCTG GTCACCAC-3' and R–*TYMP*-E7: 5'-AGCCTCTGAC CCACGTCGA-3' (PCR product: 594 bp). Then, the amplicon was sequenced with both forward and reverse primers using ABI BigDye Terminator Cycle Sequencing

Kit (Applied Biosystems®, USA). Sanger sequencing result was analyzed by NCBI BLAST (<https://blast.ncbi.nlm.nih.gov>) and CodonCode Aligner (<http://www.codoncode.com/aligner/>). Sanger sequencing confirmed this mutation as homozygous in the proband (Fig. 2).

This mutation has not been reported yet in main variant databases and our public database (Bayangene). The mutation occurred in domains described as Glycos_transf_3 in Pfam database or Nucleoside phosphorylase/phosphoribosyltransferase catalytic domain in Gene3D database which affects this domain and other domains such as PYNP_C, Pyrimidine nucleoside phosphorylase C-terminal domain, TIGR02644, THYMIDINE PHOSPHORYLASE and TP_PyNP which are described by a different database.

To reveal the conservation of amino acid sequence of *TYMP* protein across various species, multiple sequence alignment analysis by BLAST available on ExPASy (<https://web.expasy.org/cgi-bin/blast/blast.pl>) was also

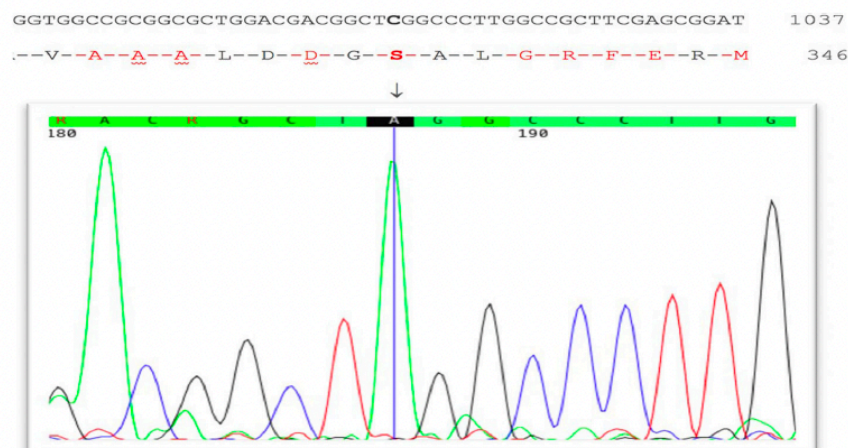


Fig. 2 Sanger sequencing electropherogram of the novel identified variant in this study. The arrow shows the position of homozygous mutation in *TYMP* (c. 1013 C > A, p.S338X)

performed. Multiple sequence alignment showed high conservation of amino acids of this protein across different species, mainly in the mutated region and after this position. (Fig. 3).

Discussion and conclusion

A group of mitochondrial disorders are characterized by mutations in the nuclear genome affecting expression and replication of the genes on the mitochondrial genetic material. Progressive external ophthalmoplegia (PEO) was the first disease identified in this group caused by defects in the intergenomic communication [19]. MNGIE, a rare progressive multisystem autosomal recessive disorder caused by a mutation in *TYMP* gene is also a member of this group of disorders.

TYMP gene which encodes the cytosolic enzyme named thymidine phosphorylase, TP, is located at chromosome 22q13.33. This protein catalyzes phosphorolysis of thymidine and deoxyuridine to 2-deoxyribose 1-phosphate and their corresponding bases (Fig. 4). TP plays both a direct and an indirect role in the metabolic pathways of various cells including those in the brain, muscle, WBCs, and bone marrow [20]. Absence of TP in cells which normally have a high expression of the enzyme (e.g., white blood cells) leads to systemic accumulation of dThd and dUrd which has a toxic effect on other tissues [21]. MNGIE usually presents with

symptoms of gastrointestinal dysfunction, such as gastrointestinal motility disorders, gastro-esophageal reflux, dysphagia, abdominal pain and distention, and diarrhea leading to severe weight loss and cachexia [22]. At this stage of the disease, most of the patients are misdiagnosed as malabsorption syndrome, inflammatory bowel disease (IBD), anorexia nervosa, or intestinal pseudo-obstruction, often leading to unnecessary medical interventions and delay in diagnosis of up to 10 years [11, 23–25]. Gastrointestinal signs and symptoms have also been observed in other genetic mitochondrial disorders as well [26]. Ptosis, ophthalmoparesis, hearing loss, and sensory-motor neuropathy constitute the most common neurologic features of patients with MNGIE [13]. Due to the high metabolic activity of extraocular muscles, deterioration in their function resulting in ophthalmoplegia or ophthalmoparesis occurs early in the course of the disease that parallels the disease progression [27]. Neuroimaging studies such as brain MRI and magnetic resonance spectroscopy (MRS) might yield a clue about the diagnosis of MNGIE, with the absence of leukoencephalopathy ruling out MNGIE in most cases. Unlike the patient reported here, who had involvement of splenium of the corpus callosum, it is relatively spared in most individuals [13, 15, 28]. Our patient was found to have leukoencephalopathy, with diffuse T₂ hyperintensity in both cerebral hemispheres white matter on brain MRI,

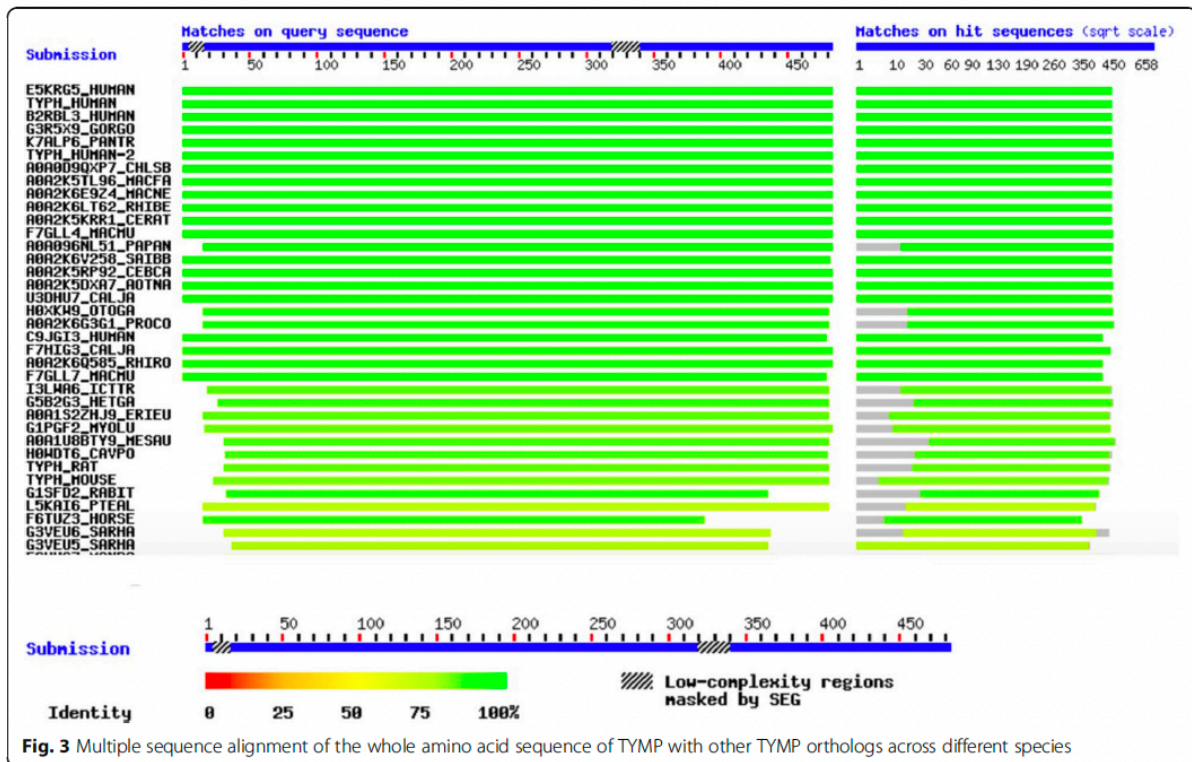


Fig. 3 Multiple sequence alignment of the whole amino acid sequence of TYMP with other TYMP orthologs across different species

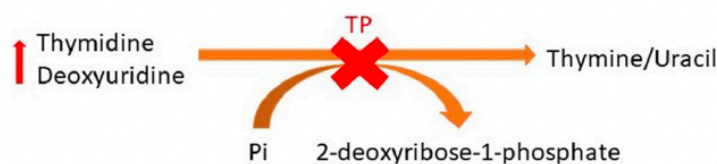


Fig. 4 The defect in the chemical reaction catalyzed by the enzyme thymidine phosphorylase in MNGIE

though, she had normal cognitive function. This, in turn, is likely to be due to the impaired blood-brain barrier function in these patients leading to edema in lieu of demyelination [29].

Other disorders with phenotypes similar to MNGIE, caused by mutations in *RRM2B* and *POLG* genes have been reported [30, 31]. Therefore, it is prudent to test the individuals suspected of having MNGIE for these genes as well. In our patient whole exome sequencing was done and no mutations in these genes were detected.

To the best of our knowledge, around 80 pathogenic mutations in the *TYMP* gene have so far been reported. Attempts to draw a genotype-phenotype correlation in this disorder have mostly been discouraging, except for c. 622G > A variant (p.Val208Met), producing less severe TP dysfunction, leading to a late-onset disease [8, 32, 33].

Current treatment modalities for MNGIE mainly focus on restoration of the activity of TP and lowering the circulatory levels of the nucleoside substrates. Hematopoietic stem cell transplantation (HSCT) has so far been used to restore TP enzyme activity in patients with MNGIE. A retrospective analysis of 24 patients who underwent HSCT for the treatment of MNGIE revealed a survival rate of 37.5% after a median follow-up of almost 4 years. Of the fifteen patients who had died, nine had died from transplant-related mortality, and six from their disease [34]. It was found that younger patients without gastrointestinal dysmotility and liver disease receiving HSCT from an HLA-matched donor would benefit mostly from this type of treatment, highlighting the importance of diagnosis in the momentous days early in the course of the illness, when HSCT would change the course of the disease [34]. Hemodialysis and peritoneal dialysis have also been proposed as treatment modalities in these patients intending to remove the nucleosides from the peripheral circulation. In a 20-year-old patient with MNGIE, peritoneal dialysis was shown to remove approximately 100 μ moles per day of thymidine and 2'-deoxyuridine from the peritoneal cavity. Although there were improvements in the symptoms, nucleosides serum levels remained unchanged [35]. A prospective study evaluating a 29-year-old patient with MNGIE who underwent extensive

hemodialysis for 1 year also revealed that it has only a transient effect on the serum and urine levels of nucleosides with no long-term effects; there were no changes in the level of the toxic metabolites in the CSF in both short-term (within 24 h) and long-term (at months 6 and 12) [36]. Our patient in this report underwent hemodialysis with only mild improvements in the gastrointestinal symptoms. These findings cast doubt on the efficacy of dialysis in the treatment of MNGIE. Other therapeutic modalities including platelet infusion, which was also performed in our patient, and orthotopic liver transplantation have also been reported for the treatment of this disease in the literature [37, 38]. Erythrocyte encapsulated thymidine phosphorylase (EE-TP) enzyme replacement therapy which has clinical trial approval is another promising treatment option [39].

The mutation found initially by WES and subsequently confirmed using Sanger sequencing is predicted to disrupt the proper function of TYMP protein since different reports have identified frameshift mutations before and after this region resulting in the impaired TYMP [7, 40, 41].

Our patient had many of the clinical, laboratory, and imaging features seen in MNGIE. The detected novel nonsense mutation in the *TYMP* gene would be of importance for genetic counseling and subsequent early diagnosis and initiation of proper therapy. On account of the wide clinical spectrum of signs and symptoms presented by patients with MNGIE, molecular diagnostic methods would be of paramount importance.

Abbreviations

MNGIE: Mitochondrial neurogastrointestinal encephalomyopathy; TP: Thymidine phosphorylase; mtDNA: Mitochondrial DNA; ANA: Anti-nuclear antibody; anti-dsDNA: anti-double stranded DNA; EMG: Electromyography; CMAP: Compound muscle action potential; WES: Whole Exome Sequencing; PEO: Progressive external ophthalmoplegia; IBD: Inflammatory bowel disease; MRS: Magnetic resonance spectroscopy; HSCT: Hematopoietic stem cell transplantation; EE-TP: Erythrocyte encapsulated thymidine phosphorylase

Acknowledgments

The authors would like to thank the family members for participating in this study.

Authors' contributions

MAF conceived and designed the study, collected, assembled and interpreted NGS data. PH, SEJ and VRO did the clinical evaluation. MS, SBJ, and MY did the genetic studies. PH and HD drafted the manuscript. HD did the bioinformatics studies. MM and HD revised the manuscript. All authors have read and approved the manuscript.

Funding

The study was supported by the NIMAD research grant (940714) awarded to MAF and Persian BayanGene research grant (2018-01-01) awarded to MAF. The funding bodies have had no role in the study design, data collection, analysis, and writing the manuscript.

Availability of data and materials

All data are available from the corresponding author on request.

Ethics approval and consent to participate

The Ethics Committee of the Persian BayanGene Research and Training Center approved the study protocol. This investigation was conducted in accordance with the ethical principles and recommendations outlined in the Declaration of Helsinki.

Consent for publication

The patient consented to the publication of the case and accompanying clinical and genetics data. Written informed consent was obtained from the patient.

Competing interests

None to declare.

Author details

¹Persian BayanGene Research and Training Center, Shiraz University of Medical Sciences, Shiraz, Iran. ²Student Research Committee, Shiraz University of Medical Sciences, Shiraz, Iran. ³Italian Institute for Genomic Medicine (IIGM), University of Turin, Turin, Italy. ⁴Shiraz Nephro-Urology Research Center, Shiraz University of Medical Sciences, Shiraz, Iran. ⁵Clinical Neurology Research Center, Shiraz University of Medical Sciences, Shiraz, Iran. ⁶Sir William Dunn School of Pathology, University of Oxford, Oxford, UK. ⁷Department of Psychiatry and Behavioral Sciences, University of Miami Miller School of Medicine, Miami, USA.

Received: 9 September 2019 Accepted: 21 April 2020

Published online: 08 May 2020

References

- Nishino I, Spinazzola A, Hirano M. Thymidine phosphorylase gene mutations in MNGIE, a human mitochondrial disorder. *Science*. 1999;283(5402):689–92.
- Nishigaki Y, Marti R, Copeland WC, Hirano M. Site-specific somatic mitochondrial DNA point mutations in patients with thymidine phosphorylase deficiency. *J Clin Invest*. 2003;111(12):1913–21.
- Nishino I, Spinazzola A, Papadimitriou A, Hammans S, Steiner I, Hahn CD, Connolly AM, Verloes A, Guimaraes J, Maillard I, et al. Mitochondrial neurogastrointestinal encephalomyopathy: an autosomal recessive disorder due to thymidine phosphorylase mutations. *Ann Neurol*. 2000;47(6):792–800.
- Ferraro P, Pontarin G, Crocco L, Fabris S, Reichard P, Bianchi V. Mitochondrial deoxynucleotide pools in quiescent fibroblasts: a possible model for mitochondrial neurogastrointestinal encephalomyopathy (MNGIE). *J Biol Chem*. 2005;280(26):24472–80.
- Pontarin G, Ferraro P, Valentino ML, Hirano M, Reichard P, Bianchi V. Mitochondrial DNA depletion and thymidine phosphate pool dynamics in a cellular model of mitochondrial neurogastrointestinal encephalomyopathy. *J Biol Chem*. 2006;281(32):22720–8.
- Gonzalez-Vioque E, Torres-Torronteras J, Andreu AL, Marti R. Limited dCTP availability accounts for mitochondrial DNA depletion in mitochondrial neurogastrointestinal encephalomyopathy (MNGIE). *PLoS Genet*. 2011;7(3):e1002035.
- Garone C, Tadesse S, Hirano M. Clinical and genetic spectrum of mitochondrial neurogastrointestinal encephalomyopathy. *Brain*. 2011;134(Pt 11):3326–32.
- Massa R, Tessa A, Margollicci M, Micheli V, Romigi A, Tozzi G, Terracciano C, Piemonte F, Bernardi G, Santorelli FM. Late-onset MNGIE without peripheral neuropathy due to incomplete loss of thymidine phosphorylase activity. *Neuromuscul Disord*. 2009;19(12):837–40.
- Pacitti D, Levene M, Garone C, Nirmalanathan N, Bax BE. Mitochondrial Neurogastrointestinal Encephalomyopathy: into the fourth decade, What We Have Learned So Far. *Front Genet*. 2018;9:669.
- Cardaioli E, Da Pozzo P, Malfatti E, Battisti C, Gallus GN, Gaudiano C, Macucci M, Malandrini A, Margollicci M, Rubegni A, et al. A second MNGIE patient without typical mitochondrial skeletal muscle involvement. *Neurol Sci*. 2010;31(4):491–4.
- Lara MC, Valentino ML, Torres-Torronteras J, Hirano M, Marti R. Mitochondrial neurogastrointestinal encephalomyopathy (MNGIE): biochemical features and therapeutic approaches. *Biosci Rep*. 2007;27(1–3):151–63.
- Okamura K, Santa T, Nagae K, Omae T. Congenital oculoskeletal myopathy with abnormal muscle and liver mitochondria. *J Neurol Sci*. 1976;27(1):79–91.
- Hirano M, Nishigaki Y, Marti R. Mitochondrial neurogastrointestinal encephalomyopathy (MNGIE): a disease of two genomes. *Neurologist*. 2004;10(1):8–17.
- Baris Z, Eminoglu T, Dalgic B, Tumer L, Hasanoglu A. Mitochondrial neurogastrointestinal encephalomyopathy (MNGIE): case report with a new mutation. *Eur J Pediatr*. 2010;169(11):1375–8.
- M H: Mitochondrial Neurogastrointestinal Encephalopathy Disease.; 2005 Apr 22 [Updated 2016 Jan 14].
- Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*. 2009;25(14):1754–60.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kerymsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–303.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
- Zeviani M, Servidei S, Gellera C, Bertini E, DiMauro S, DiDonato S. An autosomal dominant disorder with multiple deletions of mitochondrial DNA starting at the D-loop region. *Nature*. 1989;339(6222):309–11.
- Suomalainen A, Isohanni P. Mitochondrial DNA depletion syndromes—many genes, common mechanisms. *Neuromuscul Disord*. 2010;20(7):429–37.
- Valentino ML, Marti R, Tadesse S, Lopez LC, Manes JL, Lyzak J, Hahn A, Carelli V, Hirano M. Thymidine and deoxyuridine accumulate in tissues of patients with mitochondrial neurogastrointestinal encephalomyopathy (MNGIE). *FEBS Lett*. 2007;581(18):3410–4.
- Filosto M, Scarpelli M, Tonin P, Testi S, Cotelli MS, Rossi M, Salvi A, Grotto A, Vielmi V, Todeschini A, et al. Pitfalls in diagnosing mitochondrial neurogastrointestinal encephalomyopathy. *J Inher Metab Dis*. 2011;34(6):1199–203.
- Laforce R Jr, Valdmanis PN, Dupre N, Rouleau GA, Turgeon AF, Savard M. A novel TYMP mutation in a French Canadian patient with mitochondrial neurogastrointestinal encephalomyopathy. *Clin Neurol Neurosurg*. 2009;111(8):691–4.
- Monroy N, Macias Kauffer LR, Mutchinick OM. Mitochondrial neurogastrointestinal encephalomyopathy (MNGIE) in two Mexican brothers harboring a novel mutation in the ECGF1 gene. *Eur J Med Genet*. 2008;51(3):245–50.
- Taanman JW, Daras M, Albrecht J, Davie CA, Mallam EA, Muddle JR, Weatherall M, Warner TT, Schapira AH, Ginsberg L. Characterization of a novel TYMP splice site mutation associated with mitochondrial neurogastrointestinal encephalomyopathy (MNGIE). *Neuromuscul Disord*. 2009;19(2):151–4.
- Habibzadeh P, Inaloo S, Silawi M, Dastsooz H, Farazi Fard MA, Sadeghipour F, Faghihi Z, Rezaeian M, Yavarian M, Bohm J, et al. A novel TTC19 mutation in a patient with neurological, psychological, and gastrointestinal impairment. *Front Neurol*. 2019;10:944.
- Wang HF, Wang J, Wang YL, Fan JJ, Mo GL, Gong FY, Chai ZM, Zhang J, Meng HX, Li CX, et al. A novel thymidine phosphorylase mutation in a Chinese MNGIE patient. *Acta Neurol Belg*. 2017;117(1):259–67.
- Vissing J, Ravn K, Danielsen ER, Duno M, Wibrand F, Wevers RA, Schwartz M. Multiple mtDNA deletions with features of MNGIE. *Neurology*. 2002;59(6):926–9.
- Szigeti K, Sule N, Adesina AM, Armstrong DL, Saifi GM, Bonilla E, Hirano M, Lupski JR. Increased “blood-brain barrier permeability with thymidine phosphorylase deficiency. *Ann Neurol*. 2004;56(6):881–6.
- Shaibani A, Shchelochkov OA, Zhang S, Katsonis P, Lichtarge O, Wong LJ, Shinawi M. Mitochondrial neurogastrointestinal encephalopathy due to mutations in RRM2B. *Arch Neurol*. 2009;66(8):1028–32.
- Tang S, Dimberg EL, Milone M, Wong LJ. Mitochondrial neurogastrointestinal encephalomyopathy (MNGIE)-like phenotype: an



A Novel *TTC19* Mutation in a Patient With Neurological, Psychological, and Gastrointestinal Impairment

Parham Habibzadeh^{1,2}, Soroor Inaloo³, Mohammad Silawi¹, Hassan Dastsooz^{1,4}, Mohammad Ali Farazi Fard¹, Forough Sadeghipour¹, Zahra Faghihi¹, Mohaddeseh Rezaeian¹, Majid Yavarian¹, Johann Böhm⁵ and Mohammad Ali Faghihi^{1,6*}

¹ Persian BayanGene Research and Training Center, Shiraz University of Medical Sciences, Shiraz, Iran, ² Student Research Committee, Shiraz University of Medical Sciences, Shiraz, Iran, ³ Neonatal Research Center, Shiraz University of Medical Sciences, Shiraz, Iran, ⁴ Italian Institute for Genomic Medicine (IGM), University of Turin, Turin, Italy, ⁵ Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Inserm, CNRS, Université de Strasbourg, Illkirch, France, ⁶ Center for Therapeutic Innovation, Department of Psychiatry and Behavioral Sciences, University of Miami Miller School of Medicine, Miami, FL, United States

OPEN ACCESS

Edited by:

Jie Shen,
Brigham and Women's Hospital and
Harvard Medical School,
United States

Reviewed by:

Filippo M. Santorelli,
Fondazione Stella Maris (IRCCS), Italy
Georgia Xiromerisiou,
University of Thessaly, Greece

*Correspondence:

Mohammad Ali Faghihi
mfaghihi@med.miami.edu

Specialty section:

This article was submitted to
Neurogenetics,
a section of the journal
Frontiers in Neurology

Received: 05 April 2019

Accepted: 15 August 2019

Published: 04 September 2019

Citation:

Habibzadeh P, Inaloo S, Silawi M,
Dastsooz H, Farazi Fard MA,
Sadeghipour F, Faghihi Z, Rezaeian M,
Yavarian M, Böhm J and Faghihi MA
(2019) A Novel *TTC19* Mutation in a
Patient With Neurological,
Psychological, and Gastrointestinal
Impairment. *Front. Neurol.* 10:944.
doi: 10.3389/fneur.2019.00944

Mitochondrial complex III deficiency nuclear type 2 is an autosomal-recessive disorder caused by mutations in *TTC19* gene. *TTC19* is involved in the preservation of mitochondrial complex III, which is responsible for transfer of electrons from reduced coenzyme Q to cytochrome C and thus, contributes to the formation of electrochemical potential and subsequent ATP generation. Mutations in *TTC19* have been found to be associated with a wide range of neurological and psychological manifestations. Herein, we report on a 15-year-old boy born from first-degree cousin parents, who initially presented with psychiatric symptoms. He subsequently developed progressive ataxia, spastic paraparesis with involvement of caudate bodies and lentiform nuclei with cerebellar atrophy. Eventually, the patient developed gastrointestinal involvement. Using whole-exome sequencing (WES), we identified a novel homozygous frameshift mutation in the *TTC19* gene in the patient (NM_017775.3, c.581delG: p.Arg194Asnfs*16). Advanced genetic sequencing technologies developed in recent years have not only facilitated identification of novel disease genes, but also allowed revelations about novel phenotypes associated with mutations in the genes already linked with other clinical features. Our findings expanded the clinical features of *TTC19* mutation to potentially include gastrointestinal involvement. Further functional studies are needed to elucidate the underlying pathophysiological mechanisms.

Keywords: mitochondrial diseases, mitochondrial encephalomyopathies, *TTC19*, mitochondrial complex III deficiency, neurodegenerative diseases

BACKGROUND

Mitochondrial respiratory chain (MRC), consisting of five enzymatic complexes embedded in the inner mitochondrial membrane, has an important role in providing cells with ATP. Mutations in the nuclear or mitochondrial DNA affecting MRC could result in a wide array of disorders with various neurological and non-neurological presentations (1). Complex III, consisting of 11 subunits, is responsible for the transfer of electrons from reduced coenzyme Q to cytochrome C and thus, contributes to the formation of electrochemical potential ultimately leading to the production of ATP (2).

Mitochondrial complex III deficiency, nuclear type 2 (OMIM: 615157) is an autosomal-recessive disorder caused by mutations in *TTC19* gene. *TTC19* (Entrez Gene: 54902; OMIM: 613814) is involved in the preservation of complex III function by allowing turnover of Rieske protein through removal of the N-terminal proteolytic fragment of the protein (3). Mutations in *TTC19* have been shown to be linked with rapidly progressive neurological impairment (2), spinocerebellar ataxia (4–6), progressive psychosis (4), Leigh syndrome (7), developmental delay and regression in childhood (2, 8–10), bilateral cherry red spots, and failure to thrive (10). Clinical, neuroimaging, and biochemical findings in patients with pathogenic mutations in *TTC19* gene are summarized in Table 1.

Herein, we report on a 15-year-old boy presenting with psychiatric symptoms, progressive ataxia, spastic paraparesis, bilious vomiting, and constipation with a novel homozygous frameshift mutation in *TTC19*.

CASE PRESENTATION

The patient is a 15-year-old boy, who is the first and only child of consanguineous healthy parents who were first cousins. He initially presented at the age of 7 years with psychiatric symptoms including aggressive behavior and hyperactivity for which he was under treatment with methylphenidate (Ritalin), risperidone, olanzapine, and biperiden. The patient had normal psychomotor development until the age of 13 years, when he began to develop speech difficulty. Between the age of 13 and 14, he started to experience gait disturbance and difficulty walking, which progressed in the following year and made the patient wheelchair-bound. On physical examination, the patient was emaciated and cachectic; weighing 45 kg with a height of 165 cm, his body mass index (BMI) was 16.5 kg/m². Neurological examination was significant for bilateral hyperactive deep tendon reflexes, severe ataxia, tremor, horizontal nystagmus, and spasticity, which was more pronounced in the lower extremities. The patient made limited eye contact and appeared to have intellectual impairment. In addition, musculoskeletal examination was notable for *pes cavus*.

Laboratory results including serum electrolytes, plasma ammonia, liver function test, blood amino-acid analysis, and cerebrospinal fluid examination were all normal. However, his blood lactate level was elevated to 27 mg/dL (reference range: 4.5–19.8 mg/dL). Magnetic resonance imaging (MRI) of the brain showed hypersignal changes bilaterally in caudate bodies and lentiform nuclei on T₂ and fluid attenuated inversion recovery imaging (FLAIR). The lesions appeared as hypodensities in computed tomography (CT). In addition, cerebellar atrophy was detected (Figure 1).

The patient's condition deteriorated significantly in subsequent months, leading to severe cognitive impairment and mutism. In addition, the patient developed dysphagia, bilious vomiting, and constipation. Upper endoscopy revealed erythematous lesions distributed in the entire esophagus. Furthermore, gastric mucosa in the fundus, body, and antrum was hyperemic with multiple erosions. These lesions were

also seen in the bulb and the second part of duodenum. A percutaneous endoscopic gastrostomy (PEG) tube was inserted due to feeding problems.

Total genomic DNA was extracted from the patient's blood sample using QIAamp DNA Blood Mini kit (Qiagen, Germany). Subsequently, whole-exome sequencing (WES) was performed using Illumina NextSeq500 instrument. Variants with an allele frequency of more than 0.005 in gnomAD, EXAC, and our in-house database were excluded. In addition, synonymous and non-coding variants were filtered. Subsequently, considering the autosomal-recessive pattern of inheritance, homozygous, and compound heterozygous variants were analyzed (Table S1). Finally, correlation of the patient's clinical findings with phenotypes associated with the genes harboring identified genetic variations revealed that our patient was homozygous for a previously undescribed frameshift deletion mutation in *TTC19* gene (NM_017775.3: c.581delG: p.Arg194Asnfs*16). This novel mutation along with other previously reported pathogenic variants are represented in Figure 2. Sanger sequencing of exon six of the gene was also performed in the patient and his parents, using the following forward (5'-ATTACAGTTGGCTCATCACTC-3') and reverse (5'-AGATGTTGTGTGCCCCACTA-3') primers. It was confirmed that both parents were heterozygous for the mutation, and that the proband was homozygous for this mutation (Figure 3).

Written informed consent was obtained from the patient's parents. This study was conducted in accordance with ethical standards of the declaration of Helsinki.

DISCUSSION AND CONCLUSION

Mitochondrial disorders are clinically and genetically heterogeneous, greatly hindering the diagnosis of these disorders. Recent advances made in genetic technologies, allowing sequencing and in-depth investigation of the genome, has substantially improved the diagnosis of mitochondrial diseases. Autosomal-recessive cerebellar ataxias (ARCA) are a diverse group of neurodegenerative disorders characterized by movement incoordination and unsteadiness (12). A growing number of defects in biological pathways such as deficiency of DNA repair, defects in lipoprotein assembly, chaperone dysfunction, and mitochondrial defects can lead to ARCA (12). However, in almost half of the patients the genetic cause remains elusive (13). Herein, we reported a mitochondrial ARCA in a 15-year-old patient who presented with progressive ataxia, spastic paraparesis, psychiatric, and gastrointestinal symptoms with a novel frameshift deletion in *TTC19* gene. Protein-protein interaction analysis revealed an interaction of *TTC19* with many genes, including *ZFYVE26*, which has been identified as the cause of autosomal-recessive spastic paraplegia 15 (14, 15). Furthermore, animal models have highlighted the importance of *TTC19* gene. *TTC19*-null adult *Drosophila melanogaster* exhibited reduced lifespan, low fertility, adult-onset motor impairment, and abnormal optomotor function (2).

The neurological and psychiatric manifestations of the patient described here coincides well with the manifestations reported

TABLE 1 | Clinical, neuroimaging, and biochemical findings in patients with *TTG19* mutations.

References	Ghezzi et al. (2)			Nogueira et al. (4)			Atwal (7)		Morino et al. (5)	Melchionda et al. (6)	Kunii et al. (8)	Mordaunt et al. (10)	Koch et al. (9)		Conboy et al. (11)	The present study				
Sex	F	M	F	M	M	M	F	M	F	F	M	F	M	M	F	M	M			
Age at onset	5 years	10 years	5 years	43 years	27 years	12 years	15 years	34 years	1 year	31 years	19 months	25 years	6 years	Neonatal	19 months	3 years	6 years	3.5 years	7 years	
Origin	Italian	Italian	Italian	Italian	Portuguese	Portuguese	Portuguese	Portuguese	Hispanic	Japanese	Arab	Japanese	Iraqi	Turkish	Austrian	Romanian	Romanian	Kuwaiti	Iranian	
Presenting signs and symptoms	Learning disability and gait ataxia	Learning disability and gait ataxia	Regression of language and gait ataxia	Weakness of all extremities	Mood disorder and gait ataxia	Compulsive lying	Aggressive behavior	Aggressive behavior	Developmental delay and language regression	Dysarthria	Unsteady gait with frequent falls	Mood disorder and gait ataxia	Developmental delay	Developmental lactic acidosis	Global developmental delay, ataxia, dysarthria, hypotonia	Developmental delay, ataxia, regression, hypotonia	Mild developmental delay, hypotonia	Recurrent stroke-like episodes, developmental delay	Aggressive behavior and hyperactivity	
Elevated lactate (Blood/CSF)	NA	NA	NA	NA	NA	NA	NA	NA	+NA	-NA	-/-	-/-	+NA	+NA	-/-	NA	NA	NA	+NA	
Neurological findings																				
Cognitive impairment	NA	+	NA	NA	+	+	+	+	NA	+	+	+	+	+	NA	NA	NA	+	+	
Behavioral disorder	NA	NA	NA	NA	+	+	+	+	NA	NA	+	+	+	NA	NA	NA	NA	-	NA	+
Ataxia	+	+	+	+	+	+	+	+	NA	+	+	+	+	+	+	+	+	-	+	+
Dysphagia	+	+	NA	NA	+	+	+	+	NA	NA	NA	NA	+	+	+	+	+	-	+	+
Dysarthria	+	+	+	+	+	+	+	+	NA	+	+	+	+	+	+	+	+	-	+	+
Spasticity	NA	NA	NA	+	+	+	+	+	NA	NA	NA	+	-	+	+	+	+	-	NA	+
Epilepsy	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	-	+	+	+	+	-	NA	-
Hyperactive deep tendon reflexes	+	NA	NA	NA	+	+	+	+	NA	+	-	+	-	+	+	+	+	+	NA	+
Neuroimaging features	Leukoencephalopathy, hyper-intense caudate nucleus, and cerebellar atrophy	NA	Cerebellar atrophy	NA	Oligo-ponto-cerebellar atrophy and hypersignal changes in caudate putamen, cerebellar dentate nucleus, medial midbrain, and medullary olives on T ₂ -weighted sequences				Progressing T ₂ high signal lesions in putamen, caudate body, and brainstem	Cerebellar atrophy and bilateral T ₂ high intensity at inferior olives	Mid cerebellar vermis atrophy and bilateral symmetrical T ₂ high intensity lesions in lentiform nucleus with cavitated aspects on FLAIR sequence	Cerebellar atrophy and symmetric T ₂ high intensity lesions in the inferior olives and adjacent periaqueductal gray matter	Mid cerebellar and cerebral atrophy and bilateral volume loss, patchy high signal and periventricular white matter iso-intense T ₁ foci within the lentiform nuclei	Symmetrical T ₂ -weighted hyper-intensities of basal ganglia and the caudate putamen, and the mesencephalon and cerebral gray matter	Bilateral T ₂ -weighted hyper-intensities of the putamen, caudate, and cerebral atrophy	Hyper-intensities of nucleus lentiformis and caudate, putamen, and cerebral atrophy	Hyper-intensities in caput caudate nucleus and basal parts of the putamen, increased interfoliar spaces in cerebellum	Bilateral symmetrical hyper-intensities and cystic changes of putamina and the caudate nuclei	Bilateral hypersignal T ₂ -weighted changes in caudate bodies and lentiform nuclei on T ₂ and FLAIR, cerebellar atrophy	

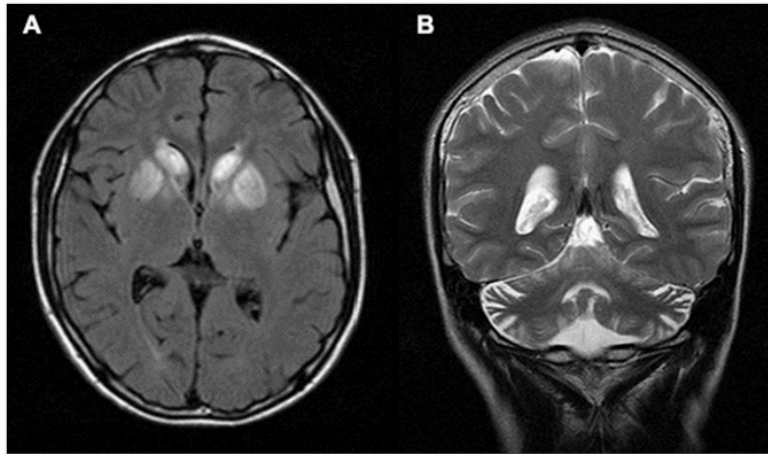


FIGURE 1 | Brain MRI. **(A)** Axial FLAIR image of the patient showing hyper-signal changes bilaterally in caudate bodies and lentiform nuclei. **(B)** Coronal T₂ image showing cerebellar atrophy.

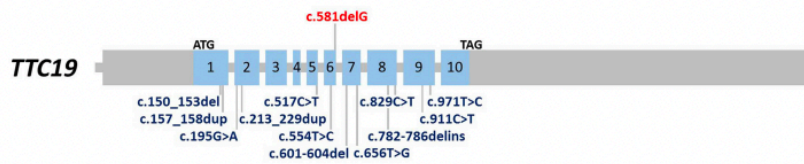


FIGURE 2 | Schematic representation of *TTC19* gene (GenBank accession no. NM_017775.3). The blue boxes indicate exons. Novel (shown in red) and previously reported mutations are shown.

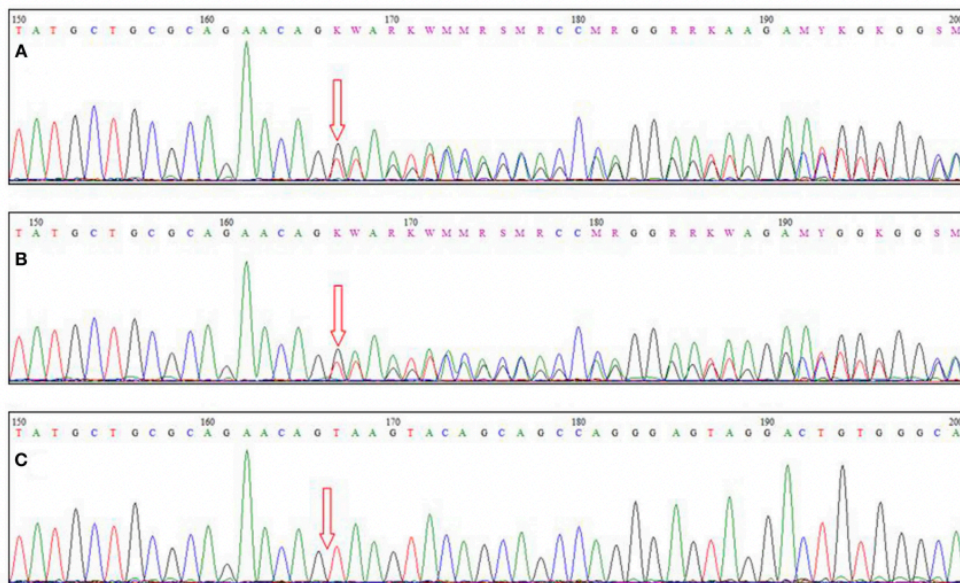


FIGURE 3 | Sequence chromatograms of **(A,B)** parents and **(C)** the patient. The arrow indicates the site of the causative mutation.

earlier; progressive signs and symptoms of basal ganglia and cerebellar dysfunction such as ataxia, dysarthria, and spasticity were observed. Furthermore, psychiatric manifestations are often seen in dominant types of cerebellar ataxia, such as spinocerebellar ataxia 7 and 8; they are generally uncommon in those with ARCA (16, 17). However, it seems that patients with mutations in *TTC19* gene are more likely to present psychiatric manifestations, as described in our patient (4, 6).

Patients with mitochondrial disorders are frequently reported to have gastrointestinal symptoms. These symptoms might be the predominant presentation in disorders like mitochondrial neurogastrointestinal encephalomyopathy (MNGIE), whereas in other disorders like Leigh syndrome, gastrointestinal symptoms are usually less prominent compared with neurological presentations (18). Despite the fact that dysphagia is reported in previous reports, constipation and bilious vomiting have not hitherto been reported in the literature. This might be due to intestinal pseudo-obstruction, which has also been reported in other mitochondrial disorders (19, 20).

Advanced genetic sequencing technologies developed in recent years, such as WES, have not only facilitated identification of novel disease genes, but also allowed revelations about novel phenotypes associated with mutations in the genes already linked with other clinical features. Our findings expanded the clinical features of *TTC19* mutation to potentially include gastrointestinal involvement. Further functional studies are however needed to shed light over the underlying pathophysiological mechanisms.

DATA AVAILABILITY

All data are available from the corresponding author on request.

REFERENCES

1. Dimauro S, Schon EA. Mitochondrial respiratory-chain diseases. *N Engl J Med.* (2003) 348:2656–68. doi: 10.1056/NEJMra022567
2. Ghezzi D, Arzuffi P, Zordan M, Da Re C, Lamperti C, Benna C, et al. Mutations in *TTC19* cause mitochondrial complex III deficiency and neurological impairment in humans and flies. *Nat Genet.* (2011) 43:259–63. doi: 10.1038/ng.761
3. Bottani E, Cerutti R, Harbour ME, Ravaglia S, Dogan SA, Giordano C, et al. *TTC19* plays a husbandry role on *UQCRC1* turnover in the biogenesis of mitochondrial respiratory complex III. *Mol Cell.* (2017) 67:96–105.e104. doi: 10.1016/j.molcel.2017.06.001
4. Nogueira C, Barros J, Sa MJ, Azevedo L, Taipa R, Torraco A, et al. Novel *TTC19* mutation in a family with severe psychiatric manifestations and complex III deficiency. *Neurogenetics.* (2013) 14:153–60. doi: 10.1007/s10048-013-0361-1
5. Morino H, Miyamoto R, Ohnishi S, Maruyama H, Kawakami H. Exome sequencing reveals a novel *TTC19* mutation in an autosomal recessive spinocerebellar ataxia patient. *BMC Neurol.* (2014) 14:5. doi: 10.1186/1471-2377-14-5
6. Kunii M, Doi H, Higashiyama Y, Kugimoto C, Ueda N, Hirata J, et al. A Japanese case of cerebellar ataxia, spastic paraparesis and deep sensory impairment associated with a novel homozygous *TTC19* mutation. *J Hum Genet.* (2015) 60:187–91. doi: 10.1038/jhg.2015.7
7. Atwal PS. Mutations in the complex III assembly factor tetratricopeptide 19 gene *TTC19* are a rare cause of leigh syndrome. *JIMD Rep.* (2014) 14:43–5. doi: 10.1007/8904_2013_282
8. Melchionda L, Damseh NS, Abu Libdeh BY, Nasca A, Elpeleg O, Zanolini A, et al. A novel mutation in *TTC19* associated with isolated complex III deficiency, cerebellar hypoplasia, and bilateral basal ganglia lesions. *Front Genet.* (2014) 5:397. doi: 10.3389/fgene.2014.00397
9. Koch J, Freisinger P, Feichtinger RG, Zimmermann FA, Rauscher C, Wagentrust HP, et al. Mutations in *TTC19*: expanding the molecular, clinical and biochemical phenotype. *Orphanet J Rare Dis.* (2015) 10:40. doi: 10.1186/s13023-015-0254-5
10. Mordaunt DA, Jolley A, Balasubramaniam S, Thorburn DR, Mountford HS, Compton AG, et al. Phenotypic variation of *TTC19*-deficient mitochondrial complex III deficiency: a case report and literature review. *Am J Med Genet A.* (2015) 167:1330–6. doi: 10.1002/ajmg.a.36968
11. Conboy E, Selcen D, Brodsky M, Gavrilova R, Ho ML. Novel homozygous variant in *TTC19* causing mitochondrial complex III deficiency with recurrent stroke-like episodes: expanding the phenotype. *Semin Pediatr Neurol.* (2018) 26:16–20. doi: 10.1016/j.spen.2018.04.003
12. Anheim M, Tranchant C, Koenig M. The autosomal recessive cerebellar ataxias. *N Engl J Med.* (2012) 366:636–46. doi: 10.1056/NEJMra1006610
13. Pandolfo M, Manto M. Cerebellar and afferent ataxias. *Continuum.* (2013) 19:1312–43. doi: 10.1212/01.CON.0000436158.39285.22
14. Hanein S, Martin E, Boukhris A, Byrne P, Goizet C, Hamri A, et al. Identification of the *SPG15* gene, encoding spastizin, as a frequent cause of complicated autosomal-recessive spastic paraplegia,

ETHICS STATEMENT

The Ethics Committee of the Persian BayanGene Research and Training Center approved the study protocol. The parents signed a written informed consent to participate in this study. Written informed consent was obtained from the parents of the patient for the publication of this case report.

AUTHOR CONTRIBUTIONS

MAF conceived and designed the study, collected, assembled, and interpreted NGS data. PH and SI clinically evaluated the patient. PH drafted the manuscript. SI, MAF, MY, and JB revised the manuscript. PH and JB did the bioinformatics analysis. MS, HD, MAFF, FS, ZF, and MR did the genetic studies.

FUNDING

This study was supported by the NIMAD research grant (940714) awarded to MAF.

ACKNOWLEDGMENTS

The authors would like to thank the family members for participating in this study.

SUPPLEMENTARY MATERIAL


The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2019.00944/full#supplementary-material>

CASE REPORT

Open Access

Clinical and molecular characterization of three patients with Hepatocerebral form of mitochondrial DNA depletion syndrome: a case series



Ghazale Mahjoub¹, Parham Habibzadeh^{1,2}, Hassan Dastsooz^{1,3}, Malihe Mirzaei¹, Arghavan Kavosi¹, Laila Jamali¹, Haniyeh Javanmardi², Pegah Katibeh⁴, Mohammad Ali Faghihi^{1,5} and Seyed Alireza Dastgheib^{6*} 

Abstract

Background: Mitochondrial DNA depletion syndromes (MDS) are clinically and phenotypically heterogeneous disorders resulting from nuclear gene mutations. The affected individuals represent a notable reduction in mitochondrial DNA (mtDNA) content, which leads to malfunction of the components of the respiratory chain. MDS is classified according to the type of affected tissue; the most common type is hepatocerebral form, which is attributed to mutations in nuclear genes such as *DGUOK* and *MPV17*. These two genes encode mitochondrial proteins and play major roles in mtDNA synthesis.

Case presentation: In this investigation patients in three families affected by hepatocerebral form of MDS who were initially diagnosed with tyrosinemia underwent full clinical evaluation. Furthermore, the causative mutations were identified using next generation sequencing and were subsequently validated using sanger sequencing. The effect of the mutations on the gene expression was also studied using real-time PCR. A pathogenic heterozygous frameshift deletion mutation in *DGUOK* gene was identified in parents of two affected patients (c.706–707 + 2 del: p.k236 fs) presenting with jaundice, impaired fetal growth, low-birth weight, and failure to thrive who died at the age of 3 and 6 months in family I. Moreover, a novel splice site mutation in *MPV17* gene (c.461 + 1G > C) was identified in a patient with jaundice, muscle weakness, and failure to thrive who died due to hepatic failure at the age of 4 months. A 5-month-old infant presenting with jaundice, dark urine, poor sucking, and feeding problems was also identified to have another novel mutation in *MPV17* gene leading to stop gain mutation (c.277C > T: p.(Gln93*)).

Conclusions: These patients had overlapping clinical features with tyrosinemia. MDS should be considered a differential diagnosis in patients presenting with signs and symptoms of tyrosinemia.

Keywords: Mitochondrial DNA depletion syndrome, *DGUOK*, *MPV17*, Mitochondrial disorders

* Correspondence: Dastgheib@sums.ac.ir

⁶Department of Medical Genetics, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran

Full list of author information is available at the end of the article



© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Background

Mitochondrial diseases are clinically and phenotypically heterogeneous disorders caused by defects in mitochondrial DNA (mtDNA) or nuclear genes encoding proteins directly or indirectly involved in mtDNA maintenance (respiratory subunits, assembly factors, enzymes, etc.). A wide range of mutations has so far been reported in patients affected with mitochondrial disorders [1–5].

MDS is inherited as autosomal recessive disorder and mainly has an early onset. It is associated with a notable reduction in mtDNA content, which causes inappropriate function of the respiratory chain components affecting a specific tissue or multiple organs such as muscle, liver, brain, and kidney [6, 7]. Regarding the affected organs, MDS is classified into four categories: myopathic, encephalomyopathic, hepatocerebral, and neurogastrointestinal forms. Each type results from different nuclear genes mutations (Additional file 1: Table S1). It has been reported that these genes have major roles in nucleotide synthesis and replication of mtDNA and their mutations disrupt mtDNA maintenance [8, 9].

Hepatocerebral type of MDS is the most common form caused by mutations in the following nuclear genes: *TWNK*, *POLG*, *DGUOK*, *MPV17*, and *TFAM* [2, 3, 10]. The hepatocerebral type generally occurs in infants of less than 6 months of age and usually results in death within the first year of life, chiefly due to hepatic failure [11].

Mutations in *DGUOK*, encoding the mitochondrial deoxyguanosine kinase, causes MDS type 3, an early-onset disease belonging to hepatocerebral form. This enzyme phosphorylates purine nucleotides to nucleotide monophosphates and provides balanced supply of nucleotides necessary for mtDNA replication [12]. *DGUOK* is a ubiquitously expressed gene with the highest expression in muscle, brain, liver, and lymphoid tissues [13].

The affected neonates are mostly diagnosed with hepatic and neurological defects, lactic acidosis, and hypoglycemia in the first few weeks of life. They present with progressive liver disease (the most common cause of death), low-birth weight, and neurological impairments (e.g., myopathy, developmental delay, nystagmus, and hypotonia) [14, 15].

MPV17, one of the recently discovered genes found to be related to hepatocerebral class, is associated with MDS type 6. It is a ubiquitously expressed gene encoding a highly conserved protein in the inner mitochondrial membrane. MDS type 6 is usually characterized by infantile- or childhood-onset progressive hepatic disease, neurological defects as well as metabolic manifestations such as lactic acidosis and hypoglycemia [16–18]. However, there have been reports of adult-onset mtDNA deletion disease due to *MPV17* mutations [19, 20]. In contrast to other forms of MDS, neurological defects are usually milder on presentation in MDS caused by mutations in *MPV17* [21].

According to some investigations, the levels of tyrosine and phenylalanine are elevated in blood or urine of those with *DGUOK* and *MPV17* mutations found in their newborn screening [22, 23]. Tyrosinemia is an inborn error of metabolism caused by impaired tyrosine metabolism [24]. Herein, we report on the disease-causing mutations in three families affected with hepatocerebral form of MDS who were initially diagnosed with tyrosinemia.

Case presentation

Whole-exome sequencing (WES)

WES was carried out on whole blood samples taken from parents of all patients to capture and enrich all exons of protein coding genes in addition to other essential parts of the genome. Next generation sequencing (NGS) was performed using Illumina HiSeq 2000 machine to sequence close to 100 million reads and standard Illumina protocol for pair-end 99 nucleotide sequencing. Basically, the test platform assayed > 95% of the target regions with sensitivity of above 99%.

Sanger sequencing

To confirm the novel mutations, whole blood samples were collected from healthy parents of all three families in EDTA tubes. All probands of the affected families had died at the time of investigation. However, we had access to the extracted DNA sample from a daughter of family II that was kept in hospital at – 20 °C, dry umbilical cord that belonged to the affected son of family III, and chorionic villus sample (CVS) from the 7-week pregnant mother of family I. DNA was extracted from all samples (blood, CVS, tissue) using QIAamp DNA Minikit (Qiagen, Germany) according to the manufacturer's instructions. The DNA concentration was then evaluated by Epoch Microplate Spectrophotometer (Bio Tek Instruments, USA) and stored at – 20 °C until use.

Following oligonucleotide PCR primers were utilized to amplify the desired genome regions: *DGUOK* (F-Exon5: 5' AAGACTGCATTGTAGCAG 3' and R-Exon5: 5' CAGCAATATTAACCTTCTGAGT 3'), *MPV17* (F-Exon4: 5' AGTGAGGTAGAGGCCTAG3' and R-Exon4: 5' CTGCACCATAACCCCTCAG 3'), *MPV17* (F-Exon7: 5' TGGTGCAGGAATGTGCTC 3' and R-Exon7: 5' CTGCAGCCTAGGTTAGAC 3').

Sanger sequencing was then conducted in both directions on the amplified DNA segments using ABI BigDye Terminator Cycle Sequencing kit (Applied Biosystems®, USA).

Real-time PCR

Total RNA was isolated from whole blood sample taken from heterozygous parents and also umbilical cord tissue of the deceased homozygous individual in family III

using Invitrogen TRIzol Reagent according to the company protocol. RNA concentration, purity and integrity was then measured by Epoch Microplate Spectrophotometer (Bio Tek Instruments, USA). The resulting RNA samples were used for cDNA synthesis using Fermentas cDNA synthesis kit (Thermo Fisher Scientific, USA). We used a Rotor-Gene Q (QIAGEN, Germany) real-time PCR cyclers by Invitrogen SYBER Green Master Mix to evaluate any alteration in *DGUOK* and *MPV17* gene expression of parents' blood and umbilical cord compared with that of a normal control. The following primer pairs were used to assess *DGUOK* and *MPV17* expression: *DGUOK*-QPCR-F: 5' TGGGAAAGTCCACGTTTGTGAA 3', *DGUOK*-QPCR-R: 5' AATGTGTAGGACCATCGTGCTG 3' and *MPV17*-QPCR-F: 5' ACTACAGCGGGATTATCCT 3', *MPV17*-QPCR-R: 5' TAACAGCAACACATTGGAC 3'. We also used glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) as the reference gene by these primer pairs: *GAPDH*-QPCR-F: 5' ACAACTTTGGTATCGTGGAAGG 3', *GAPDH*-QPCR-R: 5'GCCATCACGCCACAGTTTC3'.

Differences in the relative gene expression was evaluated by cycle threshold (Ct) values. We used $2^{-\Delta\Delta Ct}$ method to calculate relative expression between mentioned genes and *GAPDH* gene as the internal control.

Bioinformatics

Herein, several bioinformatics analyses were conducted using a wide variety of software programs. BWA aligner was used for aligning sequence reads (obtained from WES) against human genome [25]. Genome variants were identified by GATK [26]; then annotated using ANNOVAR software [27]. Public databases and standard bioinformatics software programs, such as CADD_phred, SIFT, Polyphen, Phastcons, LRT, Mutation Taster, and Mutation Assessor were used to evaluate the NGS results.

I-TASSER server (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>) was used to predict 3D structure of

proteins [28, 29]. Wild-type and mutant protein structure of proteins were then compared with UCSF chimera.

Multiple sequence alignment software program was also used to conduct comparative amino acid sequence alignment of MPV17 and DGUOK proteins.

Family I: patients I

Patient I was a 6-month-old girl, known case of liver cirrhosis, who presented with dyspnea, bloody vomiting, diarrhea, lethargy, jaundice, and dark urine. She was the second child of non-symptomatic parents who were first-degree cousins. The first child of the family presenting with similar symptoms had died at the age of 3 months. Both children had impaired fetal growth, low-birth weight, and failure to thrive. Both were floppy and hypotonic and developed prolonged jaundice and green-colored stool. The first infant weighted 1700 g at birth, had Apgar scores of 8 and 9 at the 1st and 5th min, respectively. The second child weighted 2500 g with respective Apgar scores of 8 and 8. Moreover, the first infant suffered from an unspecified ophthalmologic problem. She died at the age of 3 months due to progressive respiratory insufficiency.

On physical examination, stridor was heard over the entire lung fields on auscultation. The abdomen was distended with mild free ascitic fluid. Her vital signs included a heart rate of 142 beats/min, respiratory rate of 44 breaths/min, axillary temperature of 36.9 °C, and peripheral blood O₂ saturation of 56% on breathing the ambient air.

Laboratory evaluations indicated increased serum ALT, AST, Alk-P, AFP, and total bilirubin, and prolonged PT and PTT (Table 1). Viral markers were negative. Immunoglobulin levels were within normal range. Blood sample assay showed slightly decreased biotinidase and normal Gal-P urolytransferase activities. Tandem mass spectrometry showed elevated level of phenylalanine, tyrosine, and methionine. Based on

Table 1 Laboratory findings in the patients showing increased Alk-P, AFP, and blood tyrosine levels in all affected individuals

Variable	Patient I	Patient II	Patient III
AST (Reference range: < 60 U/L)	942	241	176
ALT (Reference range: < 45 U/L)	554	106	61
Alk-P (Reference range: 124–341 U/L)	1274	897	2700
AFP (Reference range: 0–97 ng/mL)	16,372	291	> 2000
Total Bilirubin (Reference range: < 1.9 mg/dL)	12.8	11.9	10.5
Direct Bilirubin (Reference range: < 0.2 mg/dL)	3.2	4.7	3.9
PT (Reference range: 10.5–11.5 s)	> 60	21.8	34
PTT (Reference range: 24–36 s)	58	59	58
Blood tyrosine (Reference range: 20–100 μmol/L)	240	176	309
Blood succinylacetone (Reference range: < 5.0 mcM)	< 5.0	< 5.0	< 5.0

clinical picture, the patient was initially diagnosed as having either tyrosinemia or galactosemia. On the third day of admission, the patient died. The family was referred for genetic counselling to find out whether the underlying problem was hereditary and to terminate the current pregnancy if necessary.

NGS data analysis revealed a pathogenic heterozygous frameshift deletion mutation in *DGUOK* gene (*DGUOK*: NM_001318860:exon5:c.706_707 + 2del: p.K236 fs) that was present in both parents. Homozygous or compound heterozygous *DGUOK* gene mutations have been reported in hepatocerebral MDS. Sanger sequencing confirmed the heterozygous frameshift deletion in the parents. Therefore, the disease might be inherited as an autosomal-recessive trait. Sanger sequencing study of the CVS sample revealed that the current pregnancy was a heterozygous carrier of the mutation (Fig. 1a).

Clustal Omega multiple sequencing alignment was used to align different functional isoforms of *DGUOK* proteins. The result showed that all functional isoforms of *DGUOK* protein shared the same amino acid sequence after the position 236 (Fig. 1b). This highlighted the vital role of these amino acids in the protein's functions. As mentioned above, the exact position of the observed deletion mutation was at amino acid 236 (p.K236 fs), which could be an evidence for pathogenicity

of this mutation. Mutation taster online software was also used and predicted this variation as a disease causing variant resulting in a non-sense mediated decay at position 239 in the mutant amino acid sequence [30]. 3D structure of wild-type and mutant proteins were predicted using I-TASSER server [28, 29]. UCSF chimera was used to compare these two structures (Fig. 1c). Real-time PCR analysis of *DGUOK* mRNA expression level revealed no significant difference between parents (heterozygous carriers) and healthy control (Fig. 2a).

Family II: patient II

The second patient was a 4-month-old girl who presented with jaundice since one month prior to admission. She also had weak crying, muscle weakness, poor sucking, and failure to thrive. Being a product of full-term normal vaginal delivery, she had normal APGAR score, and birth weight and head circumference. The patient had an episode of seizure when she was 12 days old. The parents were second-degree cousins and had a younger sibling who had died at the age of 9 months due to an unknown metabolic disorder. The mother also reported a previous abortion.

Detailed neurological examination revealed neurodevelopmental delay and muscle weakness in patient II. The "Fix and Follow test" of moving objects was abnormal. The physical examinations were otherwise unremarkable.

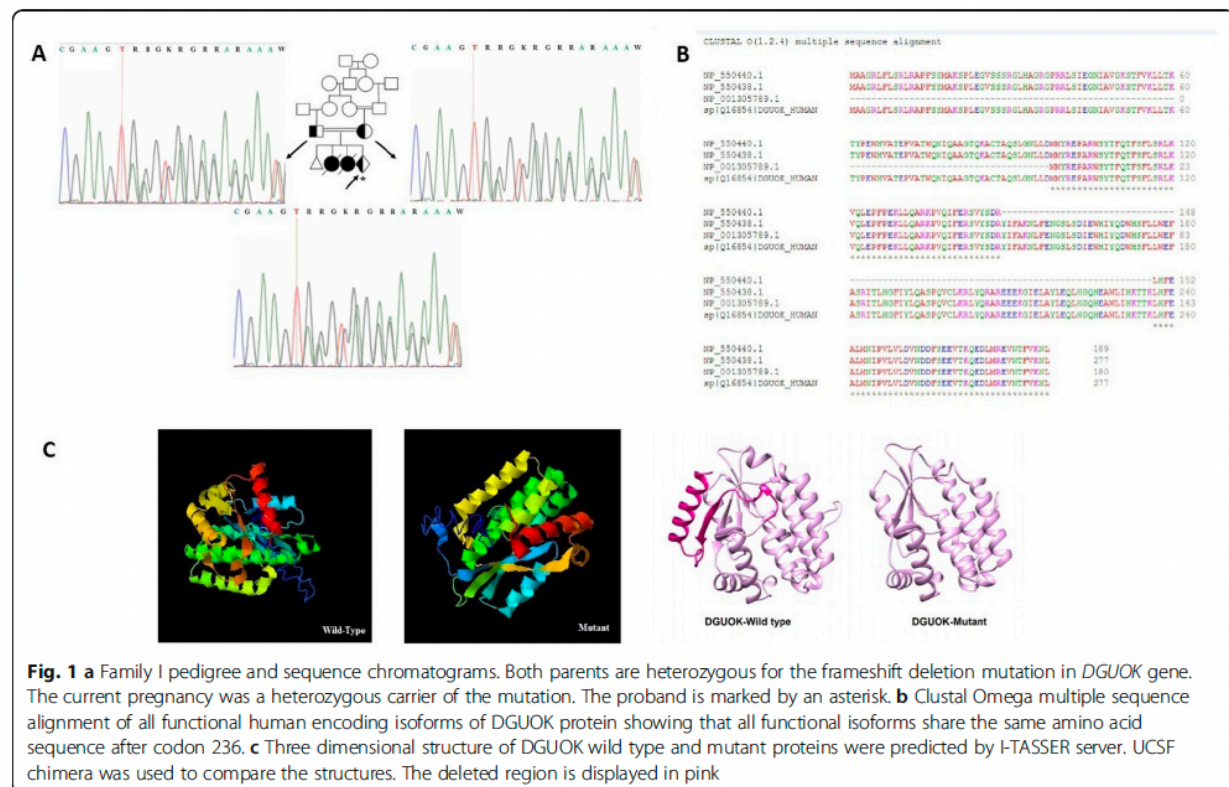
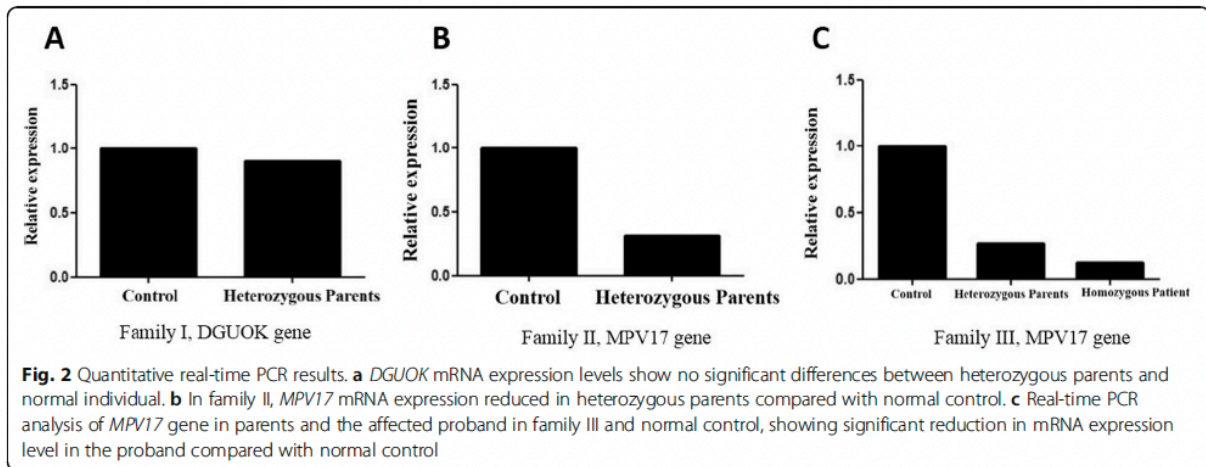


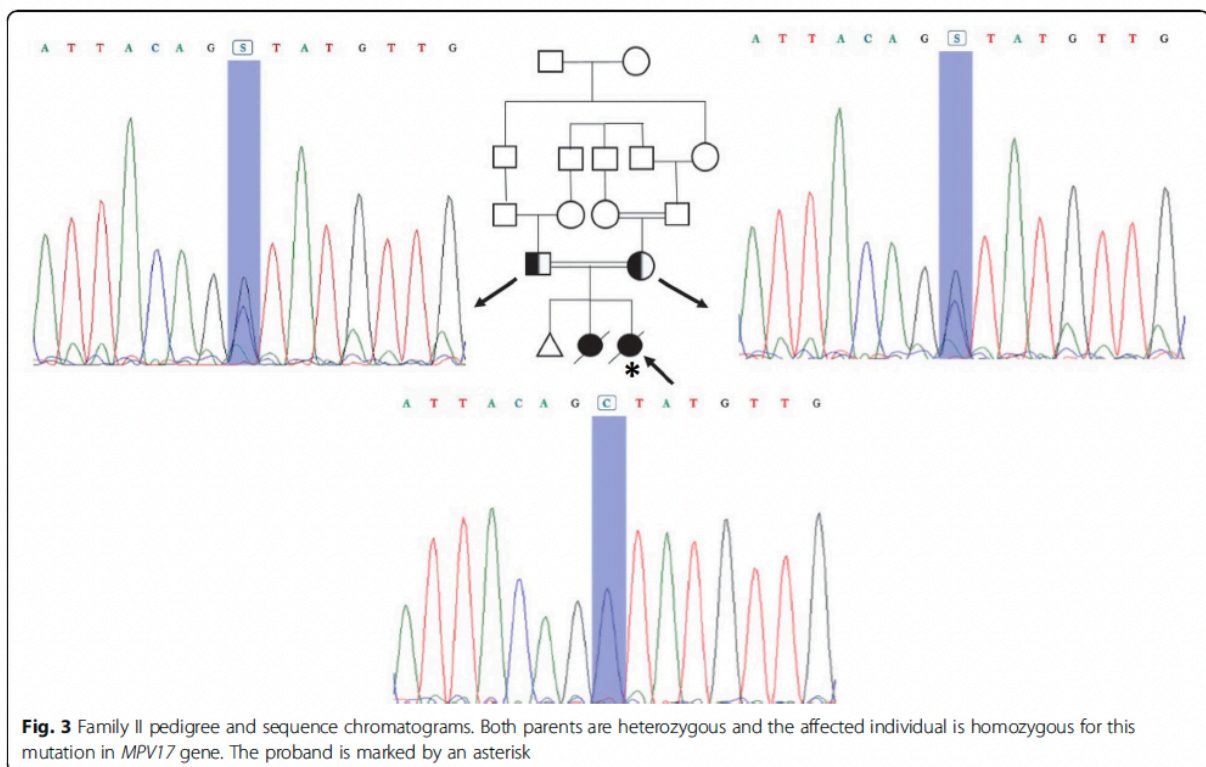
Fig. 1 a Family I pedigree and sequence chromatograms. Both parents are heterozygous for the frameshift deletion mutation in *DGUOK* gene. The current pregnancy was a heterozygous carrier of the mutation. The proband is marked by an asterisk. b Clustal Omega multiple sequence alignment of all functional human encoding isoforms of *DGUOK* protein showing that all functional isoforms share the same amino acid sequence after codon 236. c Three dimensional structure of *DGUOK* wild type and mutant proteins were predicted by I-TASSER server. UCSF chimera was used to compare the structures. The deleted region is displayed in pink



She had no abnormal findings on brain MRI. Diagnostic laboratory evaluations revealed elevated serum AST, ALT, AFP, and prolonged PT and PTT (Table 1). Tyrosine level was also elevated. Liver biopsy was in favor of cirrhosis. This patient was also initially diagnosed as having tyrosinemia. She died of hepatic failure at the age of 4 months.

NGS results showed a novel heterozygous missense (splice donor site) mutation in *MPV17* gene (*MPV17*: NM_002437:exon7:c.461 + 1G > C) in the parents. Mutation in this gene can cause autosomal-recessive

mitochondrial DNA depletion syndrome 6 (hepatocerebral type). Sanger sequencing confirmed heterozygosity and homozygosity for the mentioned splice site donor mutation in both parents and the patient, respectively, indicating the autosomal-recessive inheritance pattern for this disease (Fig. 3). According to the mutation taster, this variant is disease causing leading to change in the conserved splice site nucleotide. Conservation analysis of this nucleotide also revealed a PhyloP score of 3.966 and a Phastcons score of 1.



Analysis of *MPV17* mRNA using real-time PCR on heterozygous parents clearly indicated significant decreased level of *MPV17* mRNA expression compared with normal control. It can be concluded that one mutated copy of *MPV17* gene can affect the expression of this gene (Fig. 2b). However, one copy of this gene is sufficient to provide functional MPV17 protein in heterozygous carrier. Since we did not have RNA from the deceased child, real-time PCR only was performed on a heterozygous carrier. As a result, the mutation in *MPV17* gene described here can contribute to hepatocerebral MDS.

Family III: patient III

A 5-month-old boy, whose parents were first-degree cousins, was brought to the Pediatric Neurology ward of Namazi Hospital due to malaise, jaundice, dark urine, poor sucking, and feeding problems. He also had neurodevelopmental delay. The patient was the second child of a healthy couple and was a product of full-term normal vaginal delivery. He had normal APGAR scores. His weight and head circumference were appropriate at birth and during infancy.

He had previous history of hospital admission at the age of 16 days for prolonged jaundice, dark urine, and lethargy. Moreover, his serum ferritin level, AFP, AST, ALT, Alk-P, and total bilirubin levels were high (Table 1). All viral markers studied were negative. Ophthalmologic and gastrointestinal studies did not show any abnormalities. He also had an abnormal electroencephalogram (EEG) with epileptiform activities. Due to high serum phenylalanine and tyrosine levels in PKU screening (using tandem mass spectrometry), the patient was initially treated with PKU formula and phenobarbital. He was subsequently discharged with close clinical follow-up. He was apparently well until the age of 5 months when he was admitted with the primary diagnosis of tyrosinemia due to high tyrosine level in his serum. Otherwise, the physical examinations showed no abnormalities. Laboratory evaluations revealed increased serum AST, ALT, Alk-P, AFP, total and direct bilirubin, and tyrosine levels, and prolonged PT and PTT (Table 1). The patient died on the second day of admission.

NGS analysis results on parents revealed a novel heterozygous nonsense (stop gain) mutation in another region of *MPV17* gene (*MPV17*: NM_002437: c.277C > T: p.(Gln93^{*})). Sanger sequencing revealed that both parents were heterozygous and the proband was homozygous for this mutation (Fig. 4a). Mutation taster predicted that this variation is a disease causing variant. The comparative amino acid alignment of *MPV17* protein was conducted across different animal kingdoms using multiple sequence alignment analysis by T-Coffee Multiple Sequence

Alignment Program. The Q93 residue was highly conserved (Fig. 4b). In addition, prediction of the 3D protein structure using I-TASSER server [28, 29], revealed structural alterations resulting from this stop-gain mutation (Fig. 4c).

Real-time PCR was performed on three members of this family. mRNA expression in normal individuals was significantly higher than that in parents and the proband. It clearly demonstrated the depletion in *MPV17* expression in the proband (Fig. 2c).

Discussion and conclusions

MDS is a heterogeneous disorder that results from a reduction in mtDNA copy number. It can lead to a wide range of clinical presentations due to insufficient synthesis of the respiratory chain complexes (I, III, IV, V), which ultimately leads to insufficient energy production and mitochondrial dysfunction [6, 7, 31].

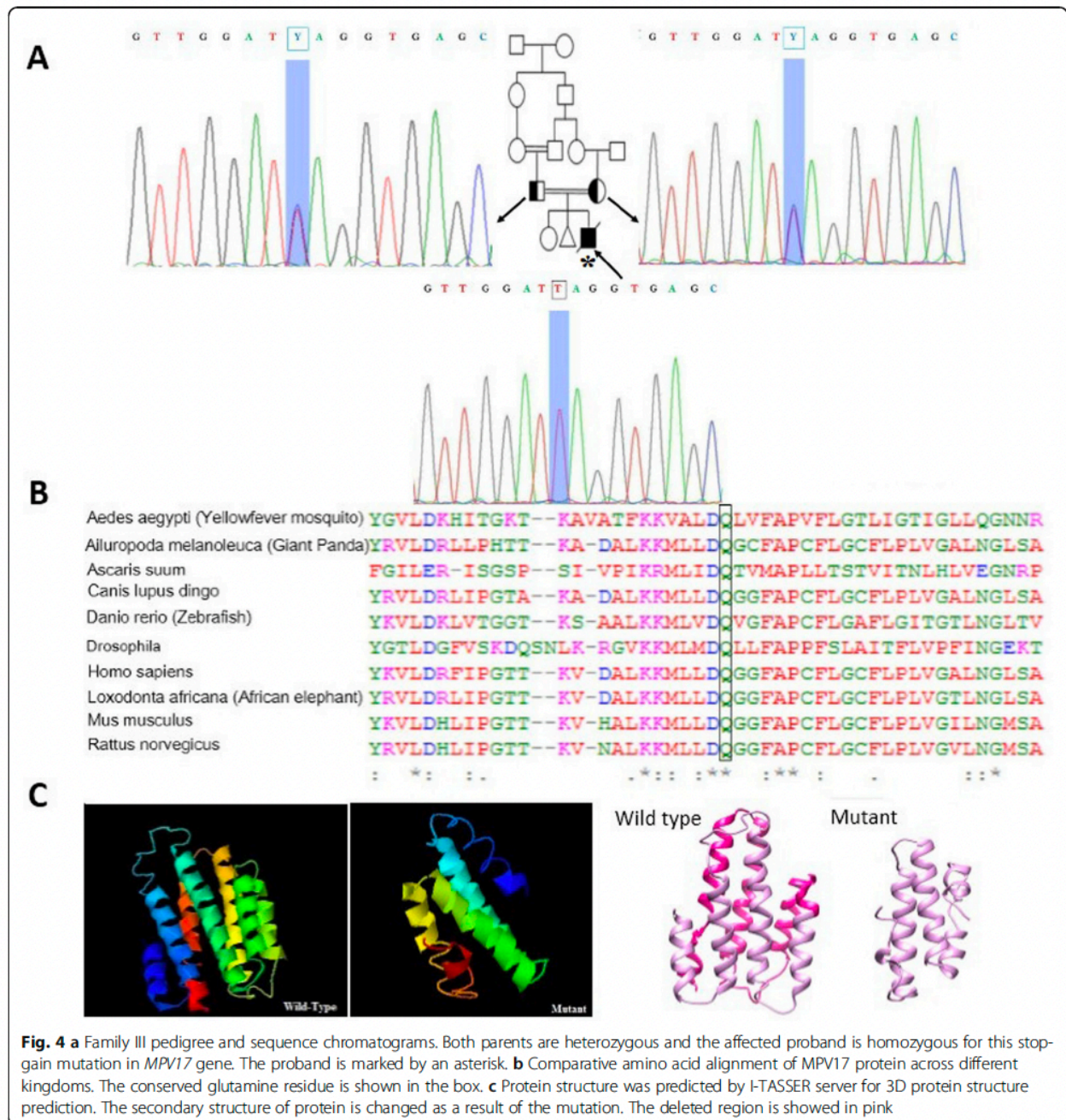
Hepatocerebral type of MDS often occurs in the infancy and its common early symptoms include persistent vomiting, failure to thrive, hypotonia, and hypoglycemia. To date, numerous genes associated with MDS have been identified. Mutations in *DGUOK* and *MPV17*, which are involved in mtDNA maintenance, have been reported in the hepatocerebral form of the disorder [2]. In this paper, we report on three hepatocerebral MDS cases which resulted from mutations of *DGUOK* gene in one family and *MPV17* gene in two families.

DGUOK has seven exons, encoding the mitochondrial deoxyguanosine kinase, which supplies dNTP for mtDNA replication. Mandel, *et al.*, illustrated a region on chromosome 2p13 by homozygosity mapping that included *DGUOK* gene in three families with hepatocerebral MDS. It was found that a nucleotide deletion (204 del A) in *DGUOK* segregated with the disease [31, 32].

With advancements in genetics sequencing techniques, especially next generation sequencing, detecting these mutations has become easier. According to The Human Gene Mutation Database (HGMD), 57 mutations have so far been reported in *DGUOK* gene—34 missense/nonsense, 6 splicing, 9 small deletions, 5 small insertions and 3 gross deletions. These mutations affect both the conserved and non-conserved *DGUOK* amino acids [33].

Whole exome sequencing of the couple in family I illustrated a deleterious heterozygous frameshift deletion mutation in *DGUOK* gene in both parents. Frameshift mutations may indirectly cause a premature termination codon and give rise to translation reading frameshift or sometimes altered splicing [34].

The parents of the patient in family I were carriers of a mutation resulting in deletion of four nucleotides (c.706–707 + 2 del AAGT) located in a splice donor site. Sezer, *et al.*, reported a 2-month-old girl with deletion mutation of *DGUOK* gene in this region (c.707 + 3–6



del TAAG) that affects splicing site and causes mitochondrial DNA depletion syndrome [35]. Here we report another patient with a deleterious mutation in this region.

In this study, the heterozygous parents showed normal *DGUOK* mRNA expression compared to normal control, reflecting that one mutated copy of this gene had no effect on mRNA expression. However, in silico investigations revealed that AAGT deletion could ultimately result in premature translation termination after codon

236, producing a truncated protein and might therefore affect the 3D protein structure or lead to non-sense mediated decay at position 239 in the mutant amino acid sequence; the wild type stop codon is located at position 278 in the amino acid sequence.

Three-dimensional protein structure of *DGUOK* was identified in 2001; two domains (β -5 and α -9) were discovered after position 245 [36]. Wang, *et al.*, showed that the C-terminal α -helix number 9 (α -9) domain of *DGUOK* has a vital role in enzyme activity as part of

phosphate donor binding site [37]. Therefore, it seems that the identified frameshift deletion in *DGUOK* gene could result in *DGUOK* deficiency and cause MDS type 3 syndrome in family I.

Patients with *DGUOK* deficiency would eventually show early-onset liver failure, which is the most common symptom. However, neurological involvement may be mild or absent [38]. Our patient had hepatomegaly and failure to thrive and died of liver failure. It is worth mentioning that the levels of three amino acids—phenylalanine, tyrosine, and methionine—were elevated in tandem mass spectrometry results of newborn screening. The observation is associated with liver failure and can be observed in large number of newborns affected by multiorgan form of the disease. Elevated hepatic enzymes concentration in the serum, direct hyperbilirubinemia, and increased gamma-glutamyltransferase levels were also observed in the affected infants, reflecting intrahepatic cholestasis [39].

Spinazzola, *et al.*, reported a new locus for hepatocerebral MDS on chromosome 2p21–23 by genome wide linkage analysis. *MPV17* was one of the top candidate genes. *MPV17* mutations segregated with the disease in all affected families [17]. The gene has eight exons, encoding a mitochondrial inner membrane protein. Although its main function is still unknown, loss of function in this protein has been shown to cause aberrant oxidative phosphorylation (OXPHOS) and mtDNA depletion in *MPV17* knock out mice model [40].

MPV17 protein, which was assumed to be located in peroxisome, has been proven to be localized in the mitochondria [17]. So far, 37 pathogenic variants have been reported in *MPV17* genes on The Human Gene Mutation Database (HGMD)—19 missense/nonsense mutations, 5 splicing variants, 6 small deletions, 2 small insertions, 1 small indels and 4 gross deletions. Most of these mutations were reported as private, except p.Arg50Gln mutation, which is the most common form [41].

In the current study, a novel missense mutation was identified in family II at a conserved splice donor site (GT at the 5' end of an intron). It causes inaccurate pattern of RNA splicing that leads to exon omission (exon skipping) or failure to splice out an intron (intron retention). Abnormal splicing may lead to a frameshift mutation at the RNA level, which induces RNA degradation or production of a truncated protein [30]. As real-time PCR results suggested, *MPV17* was significantly down-regulated in the heterozygous parents. This could be attributed to asymmetrical degradation of different splicing forms resulting from the mutation or the detrimental effect of the pathogenic variant on the regions of the gene most important for its expression.

In family III, the results showed a novel nonsense mutation (stop gain) in exon 4 of *MPV17*, which changed the p.93Q into the stop codon. The resultant mRNA might be targeted by a mechanism known as nonsense-mediated decay. The Q93 residue is highly conserved during evolution, which conveys the crucial role of glutamine in this residue. Position of the stop codon in the wild type amino acid sequence is in codon 177, while that of the mutant one is in codon 93. Therefore, the mutation would significantly affect the protein structure, which was also shown using I-TASSER server for 3D protein structure prediction. According to the Pfam database, there is a potential transmembrane helical structure might exist from codon 95 to 115 of the protein and an early stop codon in position 93 can thus disrupt the protein structure. Analysis of mRNA expression results showed that *MPV17* was significantly down-regulated in the proband compared with a normal homozygous individual. On account of the evidence presented, it can be concluded that this mutation can give rise to hepatocerebral MDS disorder.

No successful treatment has so far been found for patients with mitochondrial hepatopathy. Researchers have conducted investigations into the effect of various vitamins, cofactors and respiratory substrates for the treatment of these patients. However, none of these interventions have been found effective. Liver transplantation can only be partly effective in the absence of neurological symptoms [9].

We found one mutation in *DGUOK* gene and two novel mutations in *MPV17* in patients affected by hepatocerebral MDS. Due to increased serum tyrosine levels in most of the affected individuals, MDS has overlapping clinical features with tyrosinemia. Therefore, suspected cases should be provided with professional genetic counselling to give correct diagnosis. Moreover, physicians should be more aware of MDS as a differential diagnosis in patients with such overlapping symptoms. Molecular diagnosis may be designed to help in identifying MDS patients and families and establish accurate prenatal diagnosis of high risk individuals.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12881-019-0893-9>.

Additional file 1: Table S1. Different genes associated with mitochondrial DNA depletion syndrome.

Abbreviations

AFP: Alpha Fetoprotein; Alk-P: Alkaline phosphatase; ALT: Alanine transaminase; AST: Aspartate transaminase; CVS: Chorionic villus sampling; EEG: Electroencephalogram; MDS: Mitochondrial DNA depletion syndromes; mtDNA: Mitochondrial DNA; NGS: Next generation sequencing; PKU: Phenylketonuria; PT: Prothrombin time; PTT: Partial thromboplastin time; WES: Whole-Exome Sequencing

Acknowledgments

The authors would like to thank the family members for participating in this study.

Authors' contributions

MAF conceived and designed the study, collected, assembled, interpreted NGS data. PH, LJ, HJ, and PK clinically evaluated the patient. GM drafted the manuscript. PH, HD, SAD, and MAF revised the manuscript. GM, HD, MM, AK and SAD did the genetic studies. All authors read and approved the final manuscript.

Funding

The study was supported by the NIMAD research grant (940714) awarded to MAF. The funding body has had no role in the design of the study and collection, analysis, interpretation of data and in the writing of the manuscript.

Availability of data and materials

All data are available from the corresponding author on request.

Ethics approval and consent to participate

The Ethics Committee of the Persian BayanGene Research and Training Center approved the study protocol. The parents signed a written informed consent to participate in this study.

Consent for publication

Written informed consent for publication of the parents and patients clinical details was obtained from the parents of the patients.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Persian BayanGene Research and Training Center, Shiraz University of Medical Sciences, Shiraz, Iran. ²Student Research Committee, Shiraz University of Medical Sciences, Shiraz, Iran. ³Italian Institute for Genomic Medicine (IIGM), University of Turin, Turin, Italy. ⁴Department of Pediatrics, Shiraz University of Medical Sciences, Shiraz, Iran. ⁵Center for Therapeutic Innovation, Department of Psychiatry and Behavioral Sciences, University of Miami Miller School of Medicine, Miami, USA. ⁶Department of Medical Genetics, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran.

Received: 15 May 2019 Accepted: 10 September 2019

Published online: 29 October 2019

References

- Wallace DC. Mitochondrial diseases in man and mouse. *Science*. 1999; 283(5407):1482–8.
- Spinazzola A, Invernizzi F, Carrara F, Lamantea E, Donati A, Dirocco M, Giordano I, Meznaric-Petrusa M, Baruffini E, Ferrero I. Clinical and molecular features of mitochondrial DNA depletion syndromes. *J Inherit Metab Dis*. 2009;32(2):143–58.
- Alberio S, Mineri R, Tiranti V, Zeviani M. Depletion of mtDNA: syndromes and genes. *Mitochondrion*. 2007;7(1–2):6–12.
- Carrero-Gago L, Gamez J, Cámara Y, Alvarez de la Campa E, Aller-Alvarez JS, Moncho D, Salvado M, Galan A, de la Cruz X, Pinos T, et al. Identification and characterization of the novel point mutation m.3634A>G in the mitochondrial MT-ND1 gene associated with LHON syndrome. *Biochim Biophys Acta Mol Basis Dis*. 2017;1863(1):182–7.
- Habibzadeh P, Inaloo S, Silawi M, Dastsooz H, Farazi Fard MA, Sadeghpour F, Faghihi Z, Rezaeian M, Yavarian M, Böhm J, et al. A novel TTC19 mutation in a patient with neurological, psychological, and gastrointestinal impairment. *Front Neurol*. 2019;10:944.
- El-Hattab AW, Scaglia F. Mitochondrial DNA depletion syndromes: review and updates of genetic basis, manifestations, and therapeutic options. *Neurotherapeutics*. 2013;10(2):186–98.
- Spinazzola A, Santer R, Akman OH, Tsiakas K, Schaefer H, Ding X, Karadimas CL, Shanske S, Ganesh J, Di Mauro S. Hepatocerebral form of mitochondrial DNA depletion syndrome: novel MPV17 mutations. *Arch Neurol*. 2008;65(8):1108–13.
- Spiegel R, Saada A, Flannery PJ, Burté F, Soiferman D, Khayat M, Eisner V, Vladovski E, Taylor RW, Bindoff LA. Fatal infantile mitochondrial encephalomyopathy, hypertrophic cardiomyopathy and optic atrophy associated with a homozygous OPA1 mutation. *J Med Genet*. 2016;53(2):127–31.
- Suomalainen A, Isohanni P. Mitochondrial DNA depletion syndromes—many genes, common mechanisms. *Neuromuscul Disord*. 2010;20(7):429–37.
- Stiles AR, Simon MT, Stover A, Eftekharian S, Khanlou N, Wang HL, Magaki S, Lee H, Partynski K, Dorrani N. Mutations in TFAM, encoding mitochondrial transcription factor a, cause neonatal liver failure associated with mtDNA depletion. *Mol Genet Metab*. 2016;119(1):91–9.
- Tadiboyina VT, Rupa A, Atkison P, Feigenbaum A, Kronick J, Wang J, Hegele RA. Novel mutation in DGUOK in hepatocerebral mitochondrial DNA depletion syndrome associated with cystathioninuria. *Am J Med Genet A*. 2005;135(3):289–91.
- Copeland WC. Inherited mitochondrial diseases of DNA replication. *Annu Rev Med*. 2008;59:131–46.
- Wang L, Hellman U, Eriksson S. Cloning and expression of human mitochondrial deoxyguanosine kinase cDNA. *FEBS Lett*. 1996;390(1):39–43.
- Brahimi N, Jambou M, Sarzi E, Serre V, Boddaert N, Romano S, de Lonlay P, Slama A, Munnich A, Rötig A. The first founder DGUOK mutation associated with hepatocerebral mitochondrial DNA depletion syndrome. *Mol Genet Metab*. 2009;97(3):221–6.
- Nobre S, Grażina M, Silva F, Pinto C, Gonçalves I, Diogo L. Neonatal liver failure due to deoxyguanosine kinase deficiency. *BMJ Case Rep*. 2012;2012: bcr1220115317.
- Viscomi C, Zeviani M. MtDNA-maintenance defects: syndromes and genes. *J Inherit Metab Dis*. 2017;40(4):587–99.
- Spinazzola A, Viscomi C, Fernandez-Vizarra E, Carrara F, D'Adamo P, Calvo S, Marsano RM, Donnini C, Weiher H, Strisciuglio P. MPV17 encodes an inner mitochondrial membrane protein and is mutated in infantile hepatic mitochondrial DNA depletion. *Nat Genet*. 2006;38(5):570.
- Dalla Rosa I, Cámara Y, Durigon R, Moss CF, Vidoni S, Akman G, Hunt L, Johnson MA, Grocott S, Wang L. MPV17 loss causes deoxynucleotide insufficiency and slow DNA replication in mitochondria. *PLoS Genet*. 2016; 12(1):e1005779.
- Blakely EL, Butterworth A, Hadden RD, Bodi I, He L, McFarland R, Taylor RW. MPV17 mutation causes neuropathy and leukoencephalopathy with multiple mtDNA deletions in muscle. *Neuromuscul Disord*. 2012;22(7):587–91.
- Garone C, Rubio JC, Calvo SE, Naini A, Tanji K, DiMauro S, Mootha VK, Hirano M. MPV17 mutations causing adult-onset multisystemic disorder with multiple mitochondrial DNA deletions. *Arch Neurol*. 2012;69(12):1648–51.
- Parini R, Furlan F, Notarangelo L, Spinazzola A, Uziel G, Strisciuglio P, Concolino D, Corbetta C, Nebbia G, Menni F. Glucose metabolism and diet-based prevention of liver dysfunction in MPV17 mutant patients. *J Hepatol*. 2009;50(1):215–21.
- Uusimaa J, Evans J, Smith C, Butterworth A, Craig K, Ashley N, Liao C, Carver J, Diot A, Macleod L. Clinical, biochemical, cellular and molecular characterization of mitochondrial DNA depletion syndrome due to novel mutations in the MPV17 gene. *Eur J Hum Genet*. 2014;22(2):184.
- Dimmock D, Zhang Q, Dionisi-Vici C, Carozzo R, Shieh J, Tang LY, Truong C, Schmitt E, Sifry-Platt M, Lucio S. Clinical and molecular features of mitochondrial DNA depletion due to mutations in deoxyguanosine kinase. *Hum Mutat*. 2008;29(2):330–1.
- Endo F, Kitano A, Uehara I, Nagata N, Matsuda I, Shinka T, Kuhara T, Matsumoto I. Four-hydroxyphenylpyruvic acid oxidase deficiency with normal fumarylacetoacetase: a new variant form of hereditary hypertyrosinemia. *Pediatr Res*. 1983;17(2):92.
- Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*. 2009;25(14):1754–60.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Koryntsky A, Garimella K, Altshuler D, Gabriel S, Daly M. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–30.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010; 38(16):e164.
- Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res*. 2015;43(W1):W174–81.

CASE REPORT

Open Access

A novel frame-shift deletion in *FANCF* gene causing autosomal recessive Fanconi anemia: a case report



Soheila Zareifar¹, Hassan Dastsooz², Mahdi Shahriari³, Mohammad Ali Faghihi⁴, Golsa Shekarkhar⁵, Mohammadreza Bordbar¹, Omid Reza Zekavat¹ and Nader Shakibazad^{1,6*}

Abstract

Background: Fanconi anemia (FA) is a heterogeneous genetic disorder characterized by congenital anomalies, early-onset bone marrow failure, and a high predisposition to cancers. Up to now, different genes involved in the DNA repair pathway, mainly *FANCA* genes, have been identified to be affected in patients with FA.

Case presentation: Here, we report clinical, laboratory and genetic findings in a 3.5-year-old Iranian female patient, a product of a consanguineous marriage, who was suspicious of FA, observed with short stature, microcephaly, skin hyperpigmentation, anemia, thrombocytopenia and hypo cellular bone marrow. Therefore, Next Generation Sequencing was performed to identify the genetic cause of the disease in this patient. Results revealed a novel, private, homozygous frameshift mutation in the *FANCF* gene (NM_022725: c. 534delG, p. G178 fs) which was confirmed by Sanger sequencing in the proband.

Conclusion: Such studies may help uncover the exact pathomechanisms of this disorder and establish the genotype-phenotype correlations by identification of more mutations in this gene. It is the first report of a mutation in the *FANCF* gene in Iranian patients with Fanconi anemia. This new mutation correlates with a hematological problem (pancytopenia), short stature, and microcephaly and skin hyperpigmentation. Until now, no evidence of malignancy was detected.

Keywords: *FANCF*, Novel mutation, NGS, Autosomal recessive Fanconi Anemia

Background

Fanconi anemia (FA) is a clinically and genetically heterogeneous uncommon autosomal recessive disorder with hallmarks of congenital malformations, early-onset bone marrow failure, and a high susceptibility to malignancies due to genomic instability [1, 2]. FA is resulted from disease-causing mutations in *FANCF* genes. *FANCF* genes encode a group of proteins, which act in the pathway of DNA-damage repair along with other proteins. Up to now, 22 *FANCF* genes have been identified among which *FANCA* mutations, known as hyper-mutable

genes, have been reported to be the most common genetic causes of FA patients.

The confirmation of FA diagnosis in a proband should be considered with the following examination: Firstly, cytogenetic examination with increased levels of chromosomal breaks and radial formation on lymphocytes after exposure to Diepoxybutane (DEB) or Mitomycin C (MMC) and secondly, identification of pathogenic mutations in one of the 22 FA genes [3–5]. The aim of our study was to discover the mutated genes in an affected Iranian patient with FA using Next Generation Sequencing (NGS).

Case presentation

A 3.5-year-old girl, Caucasian, who was a product of a consanguineous marriage (first-degree cousins, Fig. 1 (timeline of case presentation) and Fig. 2a) was registered in our department due to petechia and nose

* Correspondence: nshakibazad@gmail.com

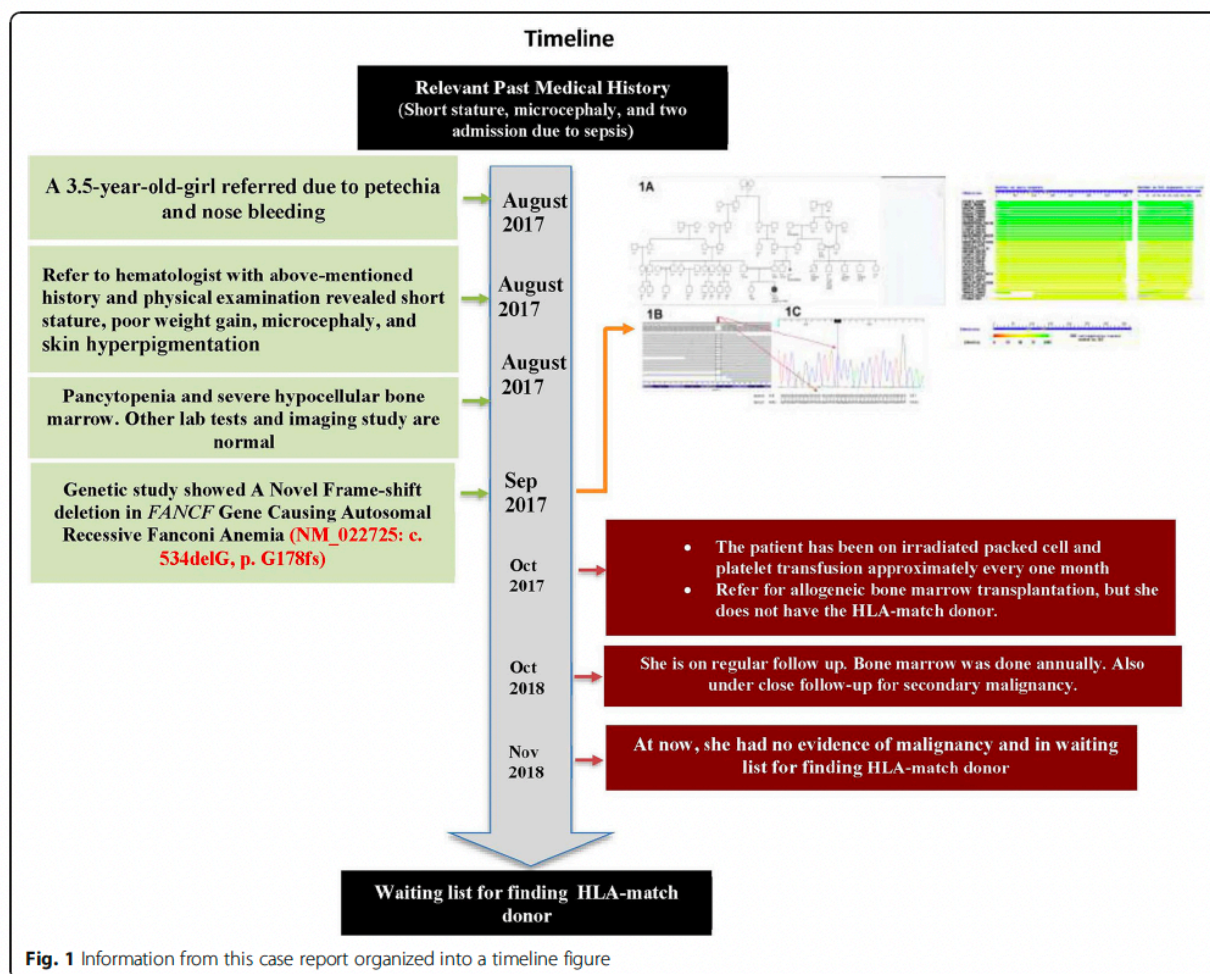
¹Hematology Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

⁶Pediatric Hematology and Oncology, Bushehr University of Medical Sciences, Bushehr, Iran

Full list of author information is available at the end of the article



© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.



bleeding. She was born at 37 weeks and 5 days of normal vaginal delivery. Her birth weight, head circumference, and height were 2.8 kg, 32 cm, and 46 cm, respectively. Although her development was in the normal range, the growth chart revealed that her height and head circumference were below the third percentile line, and her weight was around the 5th percentile line. In addition, she took zinc supplements due to her short stature. On physical examination, short stature, poor weight gain, microcephaly (head circumference; 46 cm) and skin hyperpigmentation were detected. She had a history of two-time hospital admission due to pneumonia.

Due to petechia, complete blood count was performed and the results identified anemia (Hb: 7.2 g/dL), leukopenia (WBC: $1.5 \times 10^3/\mu\text{L}$ with an absolute neutrophil count of 455), and thrombocytopenia (Platelet count: $9 \times 10^3/\mu\text{L}$). Other laboratory and imaging findings, including C3, C4, ANA, dsDNA, CH50, Immunoglobulin level, TORCH study, metabolic panel, biochemistry studies, abdomen and pelvic sonography,

brain MRI, lumbosacral and both hands X-ray, and echocardiography were normal. In addition, bone marrow aspiration and trephine biopsy revealed presented megakaryocyte and moderate to severe hypocellular bone marrow.

The patient had been on irradiated packed cell and platelet transfusion approximately every one month. She was a candidate for allogeneic bone marrow transplantation, but she did not have the HLA-match donor. She was under regular follow-up and occasionally referred due to epistaxis or pallor, and received irradiated packed cell and platelet.

Next generation sequencing (NGS)

Written informed consent was obtained from the parents. Whole blood samples were collected using EDTA tubes. Genomic DNA was prepared from peripheral leukocytes of the patient using the QIAamp DNA Blood Mini Kit (Qiagen, Germany) and then

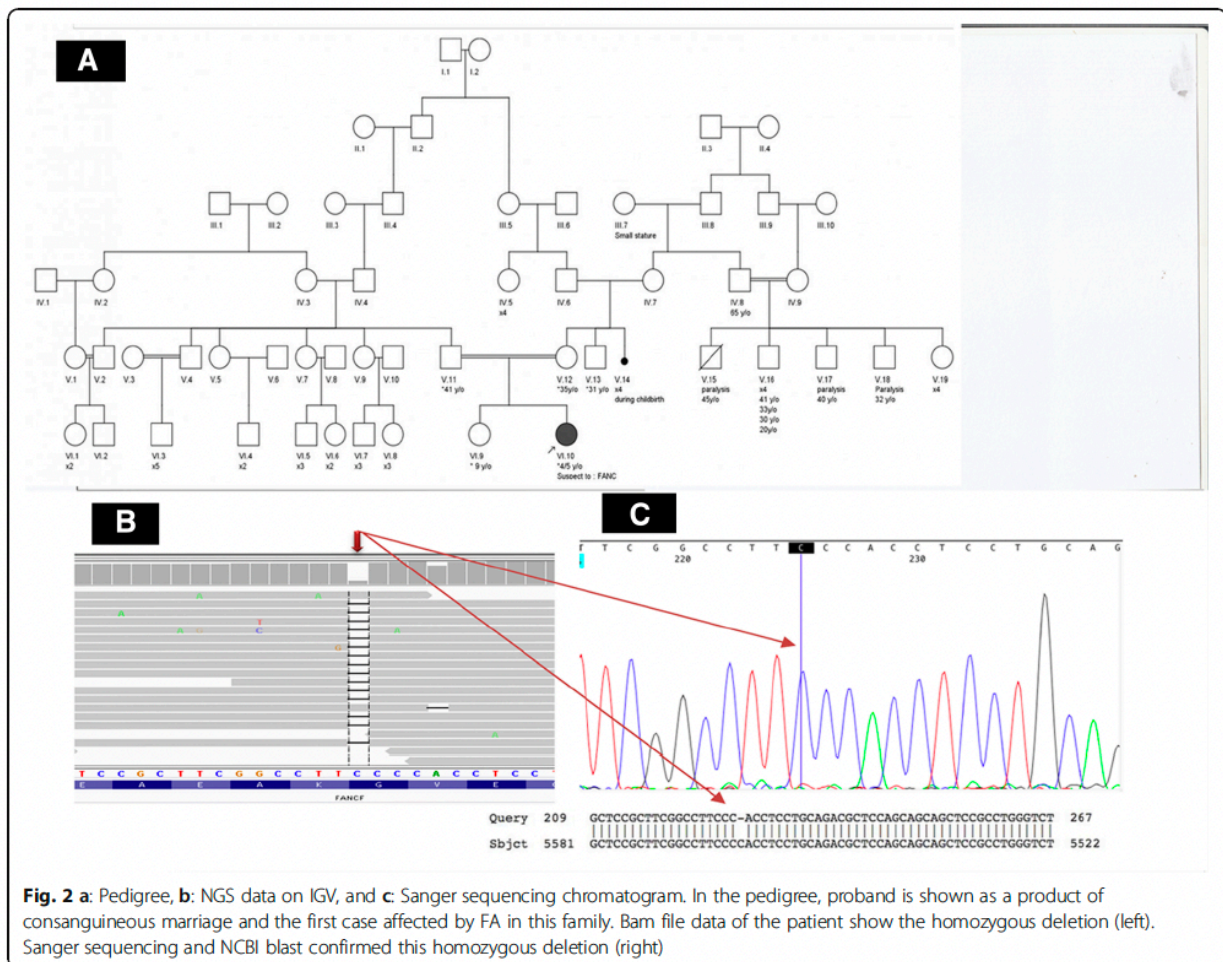


Fig. 2 a: Pedigree, **b:** NGS data on IGV, and **c:** Sanger sequencing chromatogram. In the pedigree, proband is shown as a product of consanguineous marriage and the first case affected by FA in this family. Bam file data of the patient show the homozygous deletion (left). Sanger sequencing and NCBI blast confirmed this homozygous deletion (right)

NanoDrop (ND1000, USA) was used to measure DNA concentration.

NGS covering immunological and hematological disorders was carried out on Illumina NextSeq500 machine to the sequence close to 100 million reads. Bioinformatics analysis of the sequencing results was performed using BWA aligner [6], GATK [7] and annovar [8] as well as different databases and bioinformatics software such as REVEL, MCAP, ESP6500,1000G, Clinvar, CADD-Phred, SIFT, PolyPhen, GERP, PhastCons, LRT, Mutation Assessor, Mutation Taster, phyloP46way_placental, phyloP100way_vertibrate, SiPhy_29way, FATHMM_pred, RadialSVM, ExAC, Kaviar, GME, gnomAD.

Sanger sequencing

To confirm the novel identified mutation, we performed Sanger sequencing of the genomic DNA from the proband. For this test, PCR was carried out for the patient's DNA using the following primers: F-FANCF: CGCTGGGAGATTGACATG and R-FANCF: GACCCAGTCTGTTAGCA (PCR product: 978 bp) to amplify a

mutated region of *FANCF*. Then, Sanger sequencing was used to sequence amplified DNA with both forward and reverse primers using ABI BigDye Terminator Cycle Sequencing Kit (Applied Biosystems®, USA). The analysis of Sanger sequencing data was performed with NCBI BLAST and Codon Code Aligner software. Multiple sequence alignment analysis was performed using the SIB BLAST+ Network Service From (<https://web.expasy.org/blast/>) to compare the amino acid sequence of human *FANCF* proteins with other proteins across different Kingdoms. STRING (STRING: functional protein association networks, <https://string-db.org>), tool and KEGG database (KEGG: Kyoto Encyclopedia of Genes and Genomes, <http://www.genome.jp/kegg/>) were also used to explain the FA pathway and its protein network.

Cytogenetic examination

Owing to suspicion to inherited bone marrow failure, chromosomal study with MMC on the peripheral blood lymphocyte culture of the proband was requested. To evaluate the types and rates of breakages and rearrange

ments in the chromosomes of the cells in the proband, GTG banding and the chromosome breakage test were performed on the patient's blood sample. The blood sample was then cultured and treated with different concentrations of MMC.

NGS revealed a novel, private, homozygous, frame-shift deletion mutation in the *FANCF* gene (FANCF-201, ENST00000327470.4, NM_022725: exon1, c. 534delG, p. G178 fs, position 22,625,277 on chromosome 11). Using Sanger sequencing, the mutation was confirmed in the proband as homozygous (Fig. 2b and c). The identified mutation has not been reported yet in any database of genomic variants including ESP6500, 1000 Genome Project, ExAC, Kaviar, GME, gnomAD, and our internal database (Bayan Gene), confirming the novelty of mutation. This is the first report of *FANCF* mutation in Iranian patient affected with autosomal recessive FA, complementation group F.

The comparative amino acid alignment of *FANCF* protein across most kingdoms was also carried out. As shown in Fig. 3, most of the residues were highly conserved during evolution, and any frame shift mutations could be deleterious.

In the cytogenetic study, 100 metaphase spreads were studied from cultures prepared by adding MMC and compared to age-related control. The chromosomal breakage scoring was performed on 5 different slides (one untreated with MMC, one with 150 nM MMC

concentration, one with 300 nM MMC concentration and one normal control sample treated with these 2 concentrations of MMC). 25 metaphases were evaluated on each slide for chromosomal aberration (gaps or breaks or radial formations). The results showed about 7–8 breaks/cell on average. In comparison to normal control sample which showed 0.3–0.5 breakages/cell. There was no radial formation in the normal control sample. The study showed 46, XX with multiple breaks and radial formation (quadri and triradial), compatible with Fanconi Anemia (Fig. 4, Additional file 1).

Discussion and conclusion

To date, 22 *FANCF* genes have been identified to be mutated in FA patients. The most common genes involved in FA are *BRCA2* (FA-D1, ~ 3%), *BRIP1* (FA-J, ~ 2%), *FANCA* (FA-A, 60–70%), *FANCB* (FA-B, ~ 2%), *FANCC* (FA-C, ~ 14%), *FANCD2* (FA-D2, ~ 3%), *FANCE* (FA-E, ~ 3%), *FANCF* (FA-F, ~ 2%), *FANCG* (FA-G, ~ 10%) and *FANCI* (FA-I, ~ 1%). However, less common genes are *ERCC4* (FA-Q), *FANCL* (FA-L), *FANCM* (FA-M), *MAD2L2* (FA-V), *PALB2* (FA-N), *RAD51* (FA-R), *RAD51C* (FA-O), *RFWD3* (FA-W), *SLX4* (FA-P), *UBE2T* (FA-T), *XRCC2* (FA-U) [2, 5, 9–12], *MAD2L2* (*REV7*, *FANCV*) [13], and *RFWD3* [14].

Although FA Proteins do not have similar sequences, they are correlated with their association and interactions in a common multi subunit protein

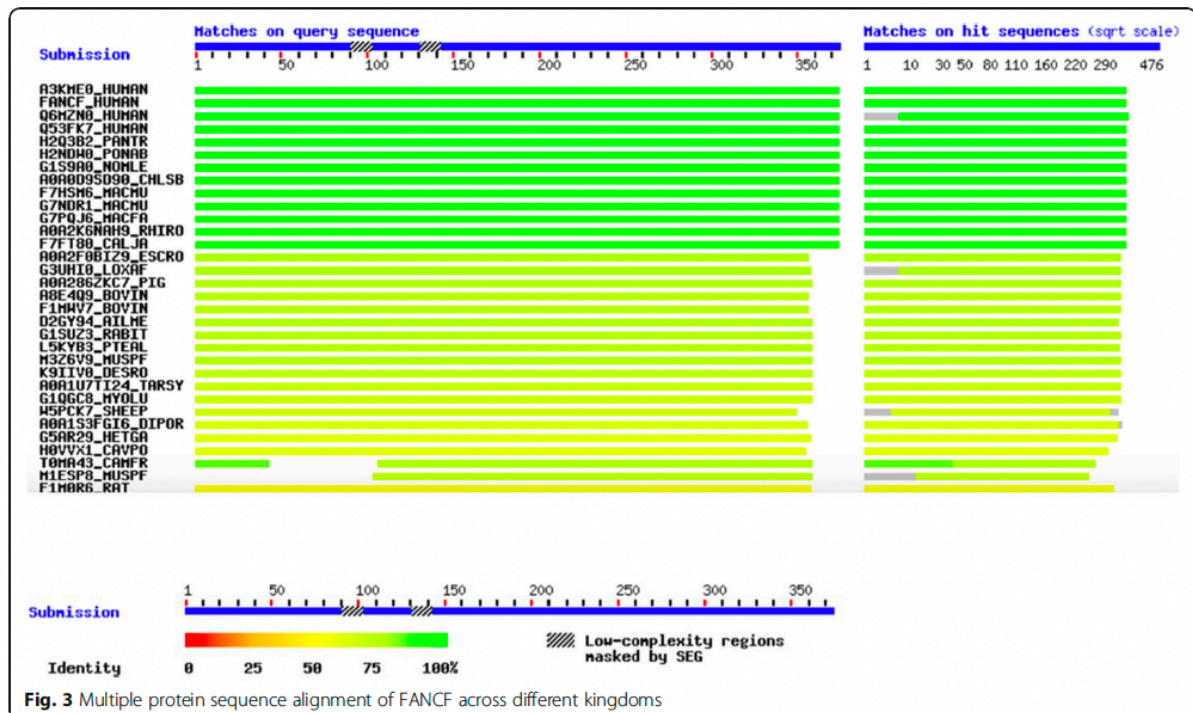
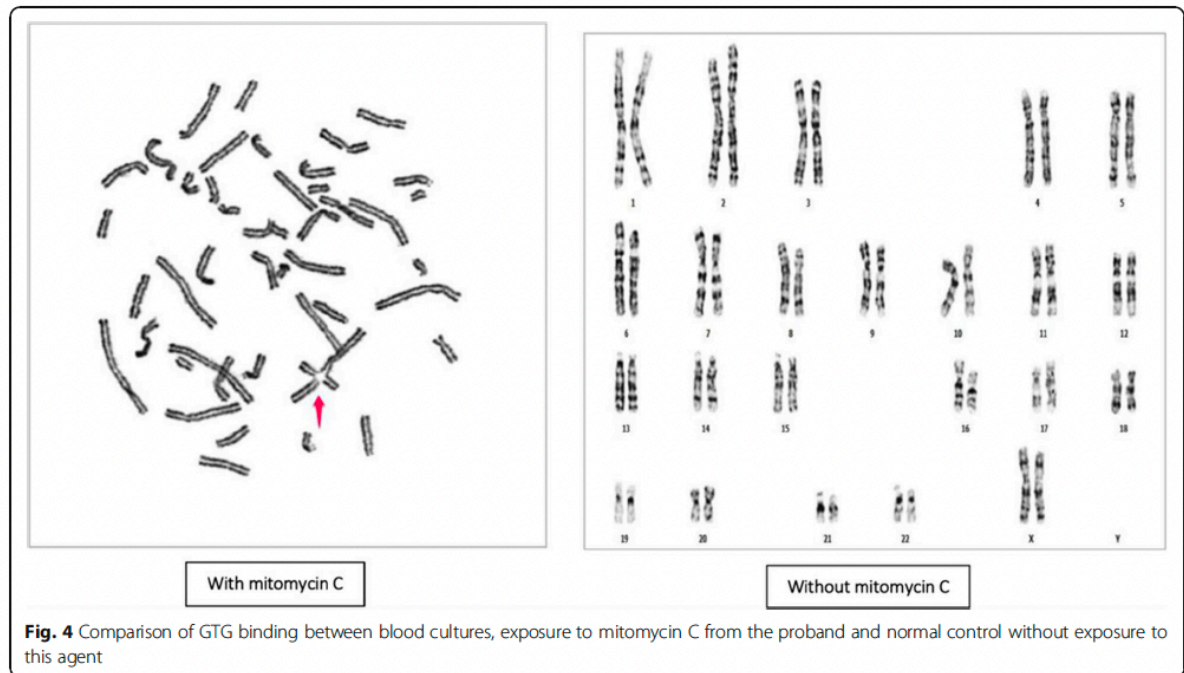
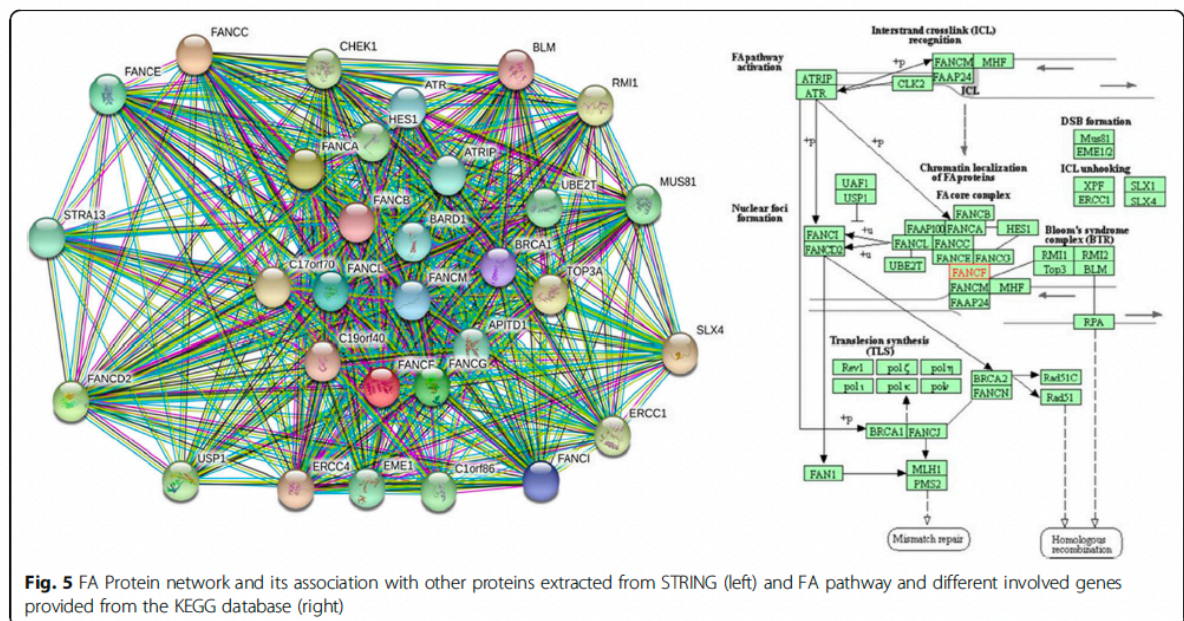


Fig. 3 Multiple protein sequence alignment of *FANCF* across different kingdoms



complex. As shown in Fig. 5, which were extracted from STRING (STRING: functional protein association networks, <https://string-db.org>) tool and KEGG database (KEGG: Kyoto Encyclopedia of Genes and Genomes, <http://www.genome.jp/kegg/>), different genes are involved in and interact with FA pathways. It can be expected to identify new genes involved in

this disorder as the list of its corresponding genes is growing. Since the FA pathway and its components play a vital role in repairing DNA damage, any impairments of these proteins result in life-threatening abnormalities and hypersensitivity to DNA cross-linking agents, leading to a high frequency of chromosomal instability [1].



In our study, we identified, for the first time, a novel, private, homozygous frame-shift deletion mutation in *FANCF* in an Iranian patient with FA. This gene, located on 11p14.3, is one of the rare genes without intron in the human genome and encodes a DNA repair protein (374 amino acids, 42,254 Da) [15]. The Pfam database indicates that amino acids 1–354 are considered as *FANCF* domain, while ProDom reported amino acids 1–355 as this domain (<https://www.ensembl.org>). Therefore, most of the amino acids are essentials for proper function of the protein.

The *FANCF* gene is the only rarely mutated in FA. At present, 14 different pathogenetic mutations of *FANCF* account for approximately 2 to 3% of the affected individuals. Disease-associated mutations have been reported throughout the single coding exon of the *FANCF* gene. The most commonly seen *FANCF* mutations are short deletions, resulting in frame-shifts and premature termination of the protein. In the study by Nicchia et al. *FANCF* loss-of-function mutation was associated with a severe phenotype characterized by multiple malformations. *FANCF* and *FANCD2* have been found to be involved in drug-resistant multiple myeloma, ovarian cancer, non-small-cell lung cancer, and head and neck cancer [16–18].

The following evidence can confirm that the identified mutation causes FA in our patient: 1- NGS only identified this mutation to be impaired in the proband and Sanger sequencing confirmed it as homozygous. 2- The novel identified mutation is a frame-shift deletion after the position of 178 in a 374 amino acid protein, coding a fully non-functional truncated protein since most of the amino acids of the *FANCF* are included in the *FANCF* domain of this protein. 3- Multiple sequence alignment revealed the conservation of more amino acids of the *FANCF* during evolution.

FANCF in complex with three *FANC* proteins, including *FANCA*, *FANCG*, and *FANCL*, interacts with *HES1* that is proposed to play a key role in the stability and nuclear localization of the FA core complex proteins. A study conducted by De Winter et al. (2000) [19] identified that *FANCF* was located primarily in the nucleus, wild-type cells, and a protein complex containing *FANCA*, *FANCC*, and *FANCG*, indicating its role in the maintenance of genomic integrity.

A report of any novel mutation in *FANCF* and other related FA genes can help shed light on the path mechanisms of this disorder and therapeutic strategies, and establishment of genotype-phenotype correlations. A novel, private, homozygous frameshift mutation in the *FANCF* gene (NM_022725: c. 534delG, p. G178 fs) in our patient correlated with hematological problem (pancytopenia), short stature, microcephaly, and skin hyperpigmentation. Until now, no evidence of malignancy was

detected. The authors of this article intend to do more family and functional study on the proband's relatives. We should also follow the patient for any further somatic mutations or incidence of any type of cancer. If this mutation is reported again in the databases we can make a group and work on genotype-phenotype correlations more practically.

Additional files

Additional file 1: (Graphical abstract): This image describe and summarize the article in picture version. (TIF 1096 kb)

Additional file 2: (Consent form): Written informed consent form was signed by the patient's father. (JPG 1497 kb)

Abbreviations

DEB: Diepoxybutane; FA: Fanconi anemia; Hb: hemoglobin; MMC: Mitomycin C; MRI: Magnetic resonance imaging; NGS: Next Generation Sequencing; PCR: Polymerase chain reaction; TORCH: Toxoplasmosis, Other (syphilis, varicella-zoster, parvovirus B19), Rubella, Cytomegalovirus, and Herpes; WBC: White blood cells

Acknowledgements

We would like to thank the patients' family for their willingness to participate in this study. We would like to thank Comprehensive Genetic Center's personnel and physicians for their help and discussion; In addition, we would like to express our deepest appreciation to Mohammad Silawi and Mohaddeseh Rezaeian for providing us the possibility to complete this report. Finally, we would like to express our gratitude to Clinical Research Development Center of Shohadaye-Khalij-e-Fars Hospital and Miss Fatemeh Gholizadeh for editorial assist and the comments they made to improve the paper.

Authors' contributions

SZ and NSH conceived and designed the study, collected, assembled, and interpreted data and wrote the manuscript. MSH, MB, and ORZ clinically evaluated the patient's interpreted data and wrote the manuscript. HD, MAF, and GSH performed the experiments, interpreted genetic data, and helped with the writing of the manuscript. All authors read and approved the final manuscript.

Funding

Nothing.

Availability of data and materials

If researchers are willing to access the data, the corresponding author should agree upon providing the interested researcher with the data.

Ethics approval and consent to participate

Ethic committee at Shiraz University of Medical Sciences, Comprehensive Medical Genetic center has approved the study and parents of the patient have signed written informed consent acknowledging their voluntary involvement in the current study.

Consent for publication

The patient's father has signed informed consent to participate in this study and to allow us to publish the result of the study (Additional file 2).

Competing interests

The authors declare that they have no competing interests.

Author details



¹Hematology Research Center, Shiraz University of Medical Sciences, Shiraz, Iran. ²Italian Institute for Genomic Medicine (IIGM), University of Turin, Turin, Italy. ³Division of Pediatric Hematology and Oncology, Department of Pediatric, Shiraz University of Medical Sciences, Shiraz, Iran. ⁴Center for Therapeutic Innovation, Department of Psychiatry and Behavioral Sciences, University of Miami Miller School of Medicine, Miami, USA. ⁵Molecular

CASE REPORT

Open Access



A novel mutation in *SEPN1* causing rigid spine muscular dystrophy 1: a Case report

Fateme Ziyaei¹, Eslam Shorafa¹, Hassan Dastsooz^{2,3}, Parham Habibzadeh^{3,4} , Hamid Nemati⁵, Amir Saeed¹, Mohammad Silawi³, Mohammad Ali Farazi Fard³, Mohammad Ali Faghihi^{3,6} and Seyed Alireza Dastgheib^{3,7*} 

Abstract

Background: Muscular dystrophies are a clinically and genetically heterogeneous group of disorders characterized by variable degrees of progressive muscle degeneration and weakness. There is a wide variability in the age of onset, symptoms and rate of progression in subtypes of these disorders. Herein, we present the results of our study conducted to identify the pathogenic genetic variation involved in our patient affected by rigid spine muscular dystrophy.

Case presentation: A 14-year-old boy, product of a first-cousin marriage, was enrolled in our study with failure to thrive, fatigue, muscular dystrophy, generalized muscular atrophy, kyphoscoliosis, and flexion contracture of the knees and elbows. Whole-exome sequencing (WES) was carried out on the DNA of the patient to investigate all coding regions and uncovered a novel, homozygous missense mutation in *SEPN1* gene (c. 1379 C > T, p.Ser460Phe). This mutation has not been reported before in different public variant databases and also our database (BayanGene), so it is classified as a variation of unknown significance (VUS). Subsequently, it was confirmed that the novel variation was homozygous in our patient and heterozygous in his parents. Different bioinformatics tools showed the damaging effects of the variant on protein. Multiple sequence alignment using BLASTP on ExPASy and WebLogo, revealed the conservation of the mutated residue.

Conclusion: We reported a novel homozygous mutation in *SEPN1* gene that expands our understanding of rigid spine muscular dystrophy. Although bioinformatics analyses of results were in favor of the pathogenicity of the mutation, functional studies are needed to establish the pathogenicity of the variant.

Keywords: Novel mutation, *SEPN1*, Rigid spine muscular dystrophy, Muscular dystrophies, Selenoproteins

Background

Muscular dystrophies are a group of disorders with heterogeneous clinical, genetic, and biochemical presentation. They are usually recognized by variable degrees of progressive muscle degeneration and weakness affecting limb, axial, and facial muscles. In some types of these disorders, muscles of the respiratory system and heart, as well as the swallowing process can be involved. Rarely, other tissues and organs, including brain, inner ear, eyes, or skin are also affected. There is a wide variability in the age of onset, symptoms and rate of progression in different forms of these disorders [1–3].

During the past decade, muscular dystrophies have extensively been studied. These advancements have been largely due to the breakthroughs developed in molecular genetics techniques, which have paved the way for the identification of the genetic and molecular basis of many of these disorders, improvements in the standards of care, and novel treatment approaches. Currently, performing molecular genetic diagnosis is very useful for establishment of phenotype-genotype correlations, pre-marital genetic counseling, prenatal diagnosis, and disease prognosis as well as identification of new treatments for these disorders [4–9].

Since identification of the exact function of genes involved in muscular dystrophies, as well as the pathological mechanisms and phenotypic consequences of mutations may shed light on therapeutic strategies for these disorders, the objective of our study was to find

* Correspondence: Dastgheib@sums.ac.ir

³Persian BayanGene Research and Training Center, Dr. Faghihi's Medical Genetic Center, Shiraz, Iran

⁷Department of Genetic, Shiraz University of Medical Sciences, Shiraz, Iran
 Full list of author information is available at the end of the article



© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

the genetic cause of muscular dystrophy in our patient and report the associated observed clinical presentations.

Case presentation

A 14-year-old boy (height = 140 cm, weight = 18 kg) from Fars province, southern Iran, who was born to first-cousin parents without family history of any genetic disorders, was referred to our center with failure to thrive, fatigue, muscular dystrophy, generalized muscular atrophy, kyphoscoliosis, and flexion contracture of the knees and elbows (Fig. 1). His motor symptoms started at the age of four years with frequent episodes of falling down that had progressed in subsequent years. There were no other family members with similar signs or symptoms. He walked at the age of 11 months and had no motor milestone delay. He had two previous admissions to the pediatric intensive care unit due to pneumonia and respiratory distress. The patient had nasal speech and sleep apnea and was under treatment with BiPAP breathing machine. By the age of 12, he was noted to have scoliosis requiring bracing.

On physical examination, the patient was cachectic with generalized muscular atrophy. Decreased muscle power in the shoulder-girdle muscles, foot extensors and limb muscles (4/5 MRC muscle scale) was noted. He also had *pes cavus* and contracture of both knees and elbows. He was also found to have severe spine rigidity with a chin-sternum distance of 15 cm.

Transthoracic echocardiography was only notable for mild pulmonary hypertension and mild tricuspid regurgitation. Pulmonary function testing revealed a FEV₁ of 35% and FVC of 32% of the predicted values.

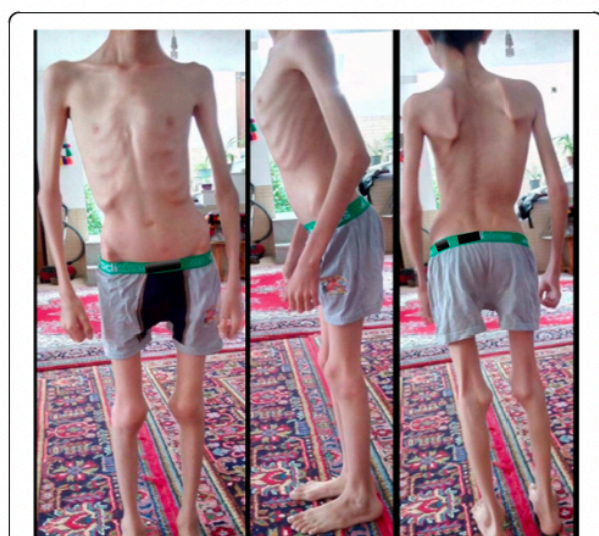


Fig. 1 Generalized muscular atrophy, kyphoscoliosis, and flexion contracture of the knees and elbows in our patient

Serum calcium and phosphorus levels were 8.2 and 3.1 mg/dL, respectively. The patient had abnormally high levels of creatine phosphokinase (CPK) (340 U/L) and lactate dehydrogenase (LDH) (1200 U/L).

Nerve conduction study was normal. However, needle electromyography (EMG) examination revealed myopathic changes in *deltoideus*, *biceps*, *tibialis anterior*, and *rectus femoris* muscles, in favor of Emery-Dreifuss muscular dystrophy.

To identify the mutated gene involved, whole-exome sequencing (WES) was used on genomic DNA extracted from EDTA blood of the patient. Next generation sequencing (NGS) was carried out on an Illumina NextSeq 500 platform to investigate all coding regions and their boundaries. WES details of coverage and number of reads are provided in Table 1. NGS data were analyzed using different bioinformatics tools and databases [10]. NGS data identified a novel, homozygous missense mutation in *SEPNI* gene (chr1:25812784, NM_020451.2: exon:10, c. 1379 C > T, p.Ser460Phe). This mutation has not been reported before in different public variant databases and also our database (BayanGene), so it is classified as a variation of unknown significance (VUS). To confirm the novel mutation identified in our patient, Sanger sequencing was performed using primers covering the mutated exon as follows:

F-SELE: 5'-GCACACACTACAGACTCAGC-3' and

R-SELE: 5'-GGAAGACACTTGGTCAAGGTTAC-3' (443 bp).

Sanger sequencing confirmed the identified mutation in *SEPNI* in proband as homozygous (T/T). His mother, father and brother were confirmed to be heterozygous (C/T), and his sister homozygous for the wildtype allele (C/C) (Fig. 2).

In order to predict the conservation of the mutated residue, we used different bioinformatics tools, including the Basic Local Alignment Search Tool (BLAST BLASTP ver 2.2.31+) on ExPASy (available from: <https://web.expasy.org/cgi-bin/blast/BLAST.pl>) and WebLogo (<https://weblogo.berkeley.edu/logo.cgi>).

Bioinformatics analysis predicted that this mutation is damaging and can affect the proper function of the protein (Table 2). In addition, multiple protein sequence alignment revealed conservation of the most residues of *SEPNI* across different species (Fig. 3a and b). Serine residue is conserved among studied species (Fig. 3a). The frequency of serine in this position is more than that for glycine and proline—only these two amino acids replace serine in other species. Since phenylalanine residue is not among these frequent amino acids, the replacement of serine with phenylalanine is predicted to be deleterious.

Discussion and conclusion

Desmin-related myopathies (DRM) are a group of muscular disorders with heterogeneous clinical presentation

Table 1 Whole Exome Sequencing detail of coverage and number of reads

Type	Value	Type	Value
Total Reads	11,709,761	Percent reads on target	95.70%
Passed filter Unique Reads aligned	11,648,030	Percent Passed filter Unique Reads aligned	99.77%
Mean Target Coverage	85X	Percent on Target	92.01%
Percent Duplicate	10.94%	Duplicate in analysis	0%
Capture Method	Agilent Inherited Disease	Total Genes Covered	3204
Run method	NextSeq 500	Sequence length	151 Pair-End
Phred Quality Score above 38	90%	GC content	55%
Nucleotide Covered GTE_1	100%	Nucleotide Covered GTE_5	99%
Nucleotide Covered GTE_8	98%	Nucleotide Covered GTE_10	97%
Nucleotide Covered GTE_15	91%	Nucleotide Covered GTE_20	83%
Nucleotide Covered GTE_30	69%	Nucleotide Covered GTE_40	56%
Nucleotide Covered GTE_50	44%	Nucleotide Covered GTE_60	35%
Nucleotide Covered GTE_70	27%	Nucleotide Covered GTE_80	21%
Nucleotide Covered GTE_90	16%	Nucleotide Covered GTE_100	13%

GTE Greater or equal to #

and genetic basis, characterized by intrasarcoplasmic desmin aggregation. It has been found that around 30% of DRM are resulted from impaired desmin gene. The impairment in the structure and subsequent accumulation of desmin result in spinal rigidity, limitation of neck and trunk flexion, progressive scoliosis, early notable limited flexion of the lumbar and cervical spine, which ultimately lead to loss of the spine and thoracic cage movements [11].

Rigid spine muscular dystrophy-1 and myopathy, congenital, with fiber-type disproportion (CFTD) result from disease-causing mutations in *SEPN1* gene (606210), located on chromosome 1p36, which encodes selenoprotein N, a glycoprotein found within the endoplasmic reticulum (ER) [12].

Selenoproteins, which contain selenocysteine residue, are vital for a wide range of biological pathways [13, 14].

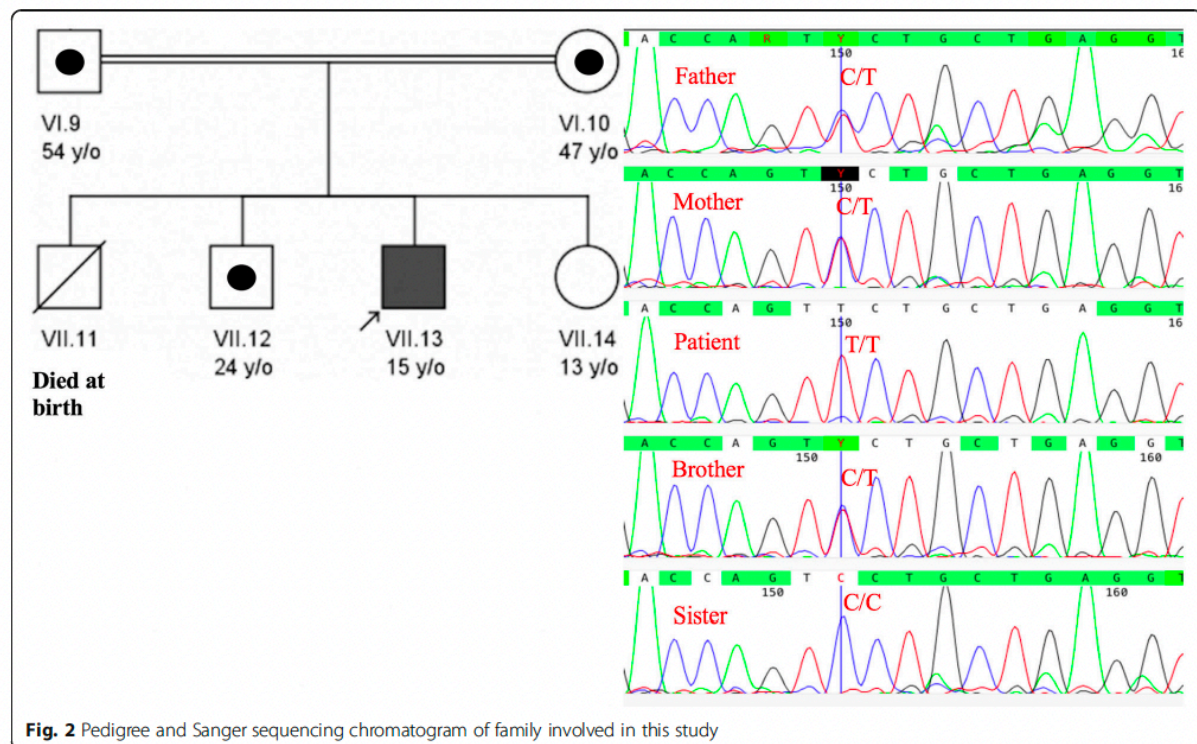
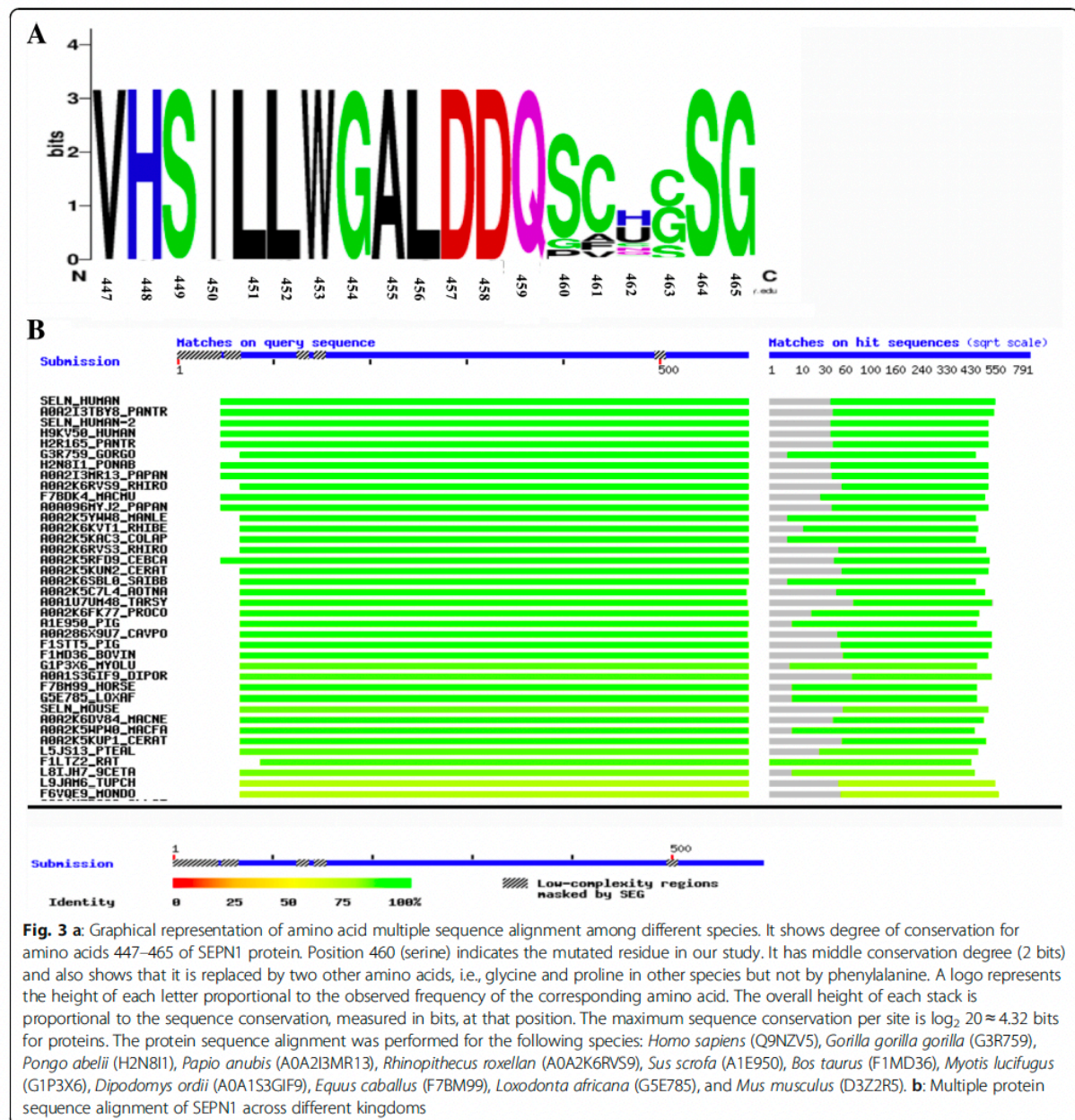


Fig. 2 Pedigree and Sanger sequencing chromatogram of family involved in this study

Table 2 Results of bioinformatics analyses of the novel mutation in this study

Chr	Start	Ref	Alt	Function	Gene	BayanGene	avsnp147	Frequency
1	25,812,784	C	T	exonic	SELENON	1	rs767530943	0.000018
SIFT	Polyphen2	LRT	Mutation Taster	Mutation Assessor	FATHMM	REVEL	MCAP	CADD_phred
D	D	N	D	M	D	0.73	.	18.91



Twenty-five selenoprotein-encoding genes have been identified in human genome [15]. The *SEPN1* contains a single selenocysteine residue and has an ER-addressing and -retention signal, indicating its localization within the ER [16, 17]. High level of expression of *SEPN1* has been found in several human fetal tissues, including muscles. However, the level of expression is lower in adult tissues, indicating its key role during early stages of embryogenesis, in early development and in cell proliferation [17, 18]. *SEPN1* has a Ca^{2+} -binding domain that is involved in the biochemical processes regulating the release of intracellular calcium. This protein is essentially involved in oxidation and reduction reactions, mainly on calcium pumps, modifying the regulation of calcium in ER [19]. Calcium homeostasis is crucial for normal development and differentiation of muscle. Therefore, *SEPN1* protein is a vital component in the process of muscle fiber formation and fiber specification [20].

In fact, supporting evidence suggests that selenium, among all its biological roles, is also influential on the normal physiologic state of striated muscles. For instance, selenium deficiency leads to acquired cardiomyopathy [21] and white muscle disease [22]. Likewise, the *SEPN1* mutations are associated with muscular dystrophies.

So far, four autosomal-recessive neuromuscular disorders, collectively regarded as *SEPN1*-related myopathies (*SEPN-RM*), have been identified [23]. Those include rigid spine muscular dystrophy (*RSMD1*) [24, 25], the classical form of multimimicore disease (*MmD*) [23], desmin-related myopathy with Mallory-body like inclusions (*MB-DRM*) [26], and *CFTD* [27]. Patients with *SEPN-RM* present with similar clinical findings, mainly early-onset hypotonia and muscular atrophy, particularly in axial musculature, along with ensuing scoliosis, neck weakness, and spinal rigidity. In patients with impaired respiratory ventilation, fatal prognosis is also expected [28]. However, the disease onset, clinical course and outcome of patients can be very variable [29].

The first *SEPN1* mutation related to a human genetic disorder was found in 2001 in patients affected by congenital *RSMD* [24]. Over the recent years, various pathogenic and non-pathogenic variations have been identified across *SEPN1* [30, 31]. The initial approach for a reliable clinical diagnosis of these disorders is doing a magnetic resonance imaging (*MRI*) of muscles, and measuring the growth hormone level, metabolic and muscular serum markers, and performing electrodiagnostic studies and echocardiography [32–38].

In conclusion, we identified a novel *SEPN1* mutation, which is predicted to be deleterious due to high damaging scores extracted from various bioinformatics software, conservation of amino acid in studied position, confirmation of mutation in the family, and absence of the mutation in our databases (1000 Iranian Genome).

Abbreviations

CFTD: Congenital fiber-type disproportion; *CPK*: Creatine phosphokinase; *DRM*: Desmin-related myopathies; *EMG*: Electromyography; *ER*: Endoplasmic reticulum; *LDH*: Lactate dehydrogenase; *MB-DRM*: Desmin-related myopathy with Mallory-body like inclusions; *MmD*: Multimimicore disease; *MRI*: Magnetic Resonance Imaging; *NGS*: Next generation sequencing; *RSMD1*: Rigid spine muscular dystrophy; *SEPN-RM*: *SEPN1*-related myopathies; *VUS*: Variation of unknown significance; *WES*: Whole exome sequencing

Acknowledgements

The authors also gratefully acknowledge the family members for participating in this research study.

Funding

This work was partly supported by the *NIMAD* research grant (940714) awarded to Mohammad Ali Faghihi. The funding agency has no role in the design of the study and collection, analysis, and interpretation of data.

Availability of data and materials

All data including *NGS* sequencing raw and analyzed data and Sanger sequencing files will be provided by corresponding author upon request. The identified mutations will be uploaded into ClinVar website.

Authors' contributions

MAF and *SAD* conceived and designed the study, collected, assembled, interpreted *NGS* data and were involved in revising the draft critically for important intellectual content. *FZ*, *ES*, *PH*, *HN* and *AS* made substantial contribution to the clinical evaluation and acquisition of data, and were involved in revising the draft critically for important intellectual content. *HD*, *MS* and *MAFF* made substantial contribution to the analysis and interpretation of data, and were involved in drafting the manuscript. All authors have given final approval of the version to be published and are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Ethics approval and consent to participate

Ethics committee at Persian BayanGene Research and Training Center, Dr. Faghihi's Medical Genetic Center has approved the study and parents of the affected individual have signed written informed consent indicating their voluntary contribution to the current study. A copy of the consent is available for review by the Editor of this journal.

Consent for publication

Both patient's legal guardians (parents) have signed informed consent to participate in this study and both families consented to publish result of study, including medical data and images.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Pediatrics, Shiraz University of Medical Sciences, Shiraz, Iran. ²Italian Institute for Genomic Medicine (IIGM), University of Turin, Turin, Italy. ³Persian BayanGene Research and Training Center, Dr. Faghihi's Medical Genetic Center, Shiraz, Iran. ⁴Student Research Committee, Shiraz University of Medical Sciences, Shiraz, Iran. ⁵Shiraz Neuroscience Research Center, Shiraz University of Medical Sciences, Shiraz, Iran. ⁶Department of Psychiatry and Behavioral Sciences, University of Miami Miller School of Medicine, Miami, USA. ⁷Department of Genetic, Shiraz University of Medical Sciences, Shiraz, Iran.

Received: 7 September 2018 Accepted: 26 December 2018

Published online: 14 January 2019

References

- Emery AE. The muscular dystrophies. *Lancet*. 2002;359(9307):687–95.
- Lisi MT, Cohn RD. Congenital muscular dystrophies: new aspects of an expanding group of disorders. *Biochim Biophys Acta*. 2007;1772(2):159–72.

Clinical values of two estrogen receptor signaling targeted lncRNAs in invasive ductal breast carcinoma

Klinické hodnoty dvou lncRNA signální dráhy estrogenového receptoru u invazivního duktálního karcinomu prsu

Ilbeigi S.¹, Naeimzadeh Y.², Davoodabadi Farahani M.¹, Rafiee Monjezi M.³, Dastsooz H.⁴, Daraei A.⁵, Farahani F.¹, Dastgheib A.¹, Mansoori Y.⁶, Bagher Tabei S. M.¹

¹ Department of Medical Genetics, School of Medicine Shiraz University of Medical Science, Shiraz, Iran

² Department of Biology, Faculty of Sciences, Yazd University, Yazd, Iran

³ Department of Medical Immunology, Shiraz University of Medical Science, Shiraz, Iran

⁴ Department of life science and system biology, University of Turin, Turin, Italy

⁵ Department of Genetics, Faculty of Medicine, Babol University of Medical Sciences, Babol, Iran

⁶ Department of Medical Genetics, Fasa University of Medical Science, Fasa, Iran

Summary

Background: Invasive ductal carcinoma (IDC) is the most frequent type of breast cancer (BC) in women, with a high clinical burden due to its high invasive properties. Despite of quickly emerging new data regarding the molecular heterogeneity of invasive cancers, far less is known about the molecular patterns among cases of IDC. An expanding body of evidence has demonstrated that dysregulation of long noncoding RNAs (lncRNAs) is involved in the heterogeneity feature of BC. **Methods:** In this study, we analyzed the expression levels of two novel lncRNAs LOC100288637 and RP11-48B3 in 51 IDC tissues in comparison with adjacent non-cancerous tissues. And finally, bioinformatic evaluation has been done. **Results:** The results of quantitative polymerase chain reaction showed that LOC100288637 and RP11-48B3 were significantly overexpressed in tumor tissues compared to normal samples ($P = 0.0085$ and $P = 0.0002$, respectively). Also, the two lncRNAs were overexpressed in both MDA-MB-231 and MCF-7 BC cell lines, nevertheless, with a higher expression pattern in MDA-MB-231 than MCF7 cell line. Furthermore, LOC100288637 had an elevated expression level in HER-2 positive tumors compared to HER-2 negative tumors ($P = 0.031$). Interestingly, the lncRNA RP11-48B3.4 was upregulated in IDC subjects with the age at menarche < 14 years compared to patients with the age at menarche ≥ 14 ($P = 0.041$). It was observed in another result that lncRNA RP11-48B3.4 is significantly upregulated in tumors with a lower histological grade compared to tumor samples with higher grades ($P = 0.047$). And finally, using bioinformatic evaluation, we found a predicted interaction between RP11-48B3.4 and mRNA zinc finger and BTB domain containing 10 (ZBTB10). **Conclusion:** Altogether, our findings suggest that these lncRNAs with potential oncogenic roles are involved in the pathogenesis of IDC with clinical significance and they may therefore serve as novel markers for the diagnosis and treatment of IDC.

Key words

invasive ductal breast carcinoma – prognosis – long noncoding RNAs – LOC100288637 – RP11-48B3

The authors declare they have no potential conflicts of interest concerning drugs, products, or services used in the study.

Autoři deklarují, že v souvislosti s předmětem studie nemají žádné komerční zájmy.

The Editorial Board declares that the manuscript met the ICMJE recommendation for biomedical papers.

Redakční rada potvrzuje, že rukopis práce splnil ICMJE kritéria pro publikace zasílané do biomedicínských časopisů.



Dr Seyed Mohammad Bagher Tabei
Department of Medical Genetics
Faculty of Medicine
Shiraz University of Medical Sciences
Zand St, Shiraz
Iran, post code: 7134845794

and

Dr Yaser Mansoori
Fasa University of Medical Sciences
Ave Sina Square
Fasa, Fars
Iran, post code: 7461686688

Submitted/Obdrženo: 31. 12. 2020

Accepted/Přijato: 25. 4. 2021

doi: 10.48095/ccko2021382

Souhrn

Východiska: Invazivní duktální karcinom (invasive ductal carcinoma – IDC) je nejčastějším typem karcinomu prsu (breast cancer – BC) u žen s vysokou klinickou zátěží v důsledku svých vysoce invazivních vlastností. I přes to, že se rychle objevují nová data týkající se molekulární heterogeneity invazivních karcinomů, mnohem méně je známo o molekulárních vzorcích IDC. Stále se zvyšující množství důkazů ukázalo, že heterogenní povaha BC souvisí s dysregulací dlouhých nekódujících RNA (long noncoding RNAs – lncRNAs). **Metody:** V této studii jsme analyzovali hladinu exprese dvou nových lncRNAs, LOC100288637 a RP11-48B3, v 51 tkáních IDC, které jsme porovnali s přílehlými nekancerózními tkáněmi. Nakonec bylo provedeno bioinformatické hodnocení. **Výsledky:** Výsledky kvantitativní polymerázové řetězové reakce ukázaly, že LOC100288637 a RP11-48B3 byly v nádorové tkáni významně overexprimovány, a to v porovnání s normálními vzorky ($p = 0,0085$ a $p = 0,0002$). Tyto dvě lncRNAs byly také overexprimovány jak v buněčné linii MDA-MB-231 tak v MCF-7 BC, ovšem v buněčné linii MDA-MB-231 byl pozorován vyšší vzorec exprese než u MCF7. Navíc LOC100288637 měl zvýšenou hladinu exprese u HER-2 pozitivních nádorů v porovnání s HER-2 negativními nádory ($p = 0,031$). Je zajímavé, že lncRNA RP11-48B3.4 byla upregulována u subjektů s IDC a menarče ve věku < 14 let, a to v porovnání s pacientkami s menarče ve věku ≥ 14 let ($p = 0,041$). Při jiném výsledku bylo pozorováno, že lncRNA RP11-48B3.4 je významně upregulována u nádorů nižšího histologického grade v porovnání se vzorky nádorů s vyšším grade ($p = 0,047$). A nakonec jsme pomocí bioinformatického hodnocení našli předpokládanou interakci mezi RP11-48B3.4 a mRNA „zinc finger and BTB domain containing 10“. **Závěr:** Naše zjištění svědčí o tom, že tyto lncRNAs, které potenciálně hrají onkogenní roli, jsou klinicky významně zapojeny do patogeneze IDC a mohou tedy sloužit jako nové markery pro diagnostiku a léčbu IDC.

Klíčová slova

invazivní duktální karcinom prsu – prognóza – dlouhé nekódující RNA – LOC100288637 – RP11-48B3

Introduction

Breast cancer (BC) is the second cause of death among women worldwide [1]. BC is a highly heterogeneous disease which is characterized by different phenotypic features, diverse response to existing therapy and unpredictable clinical outcome [2]. Invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC) are the most common variants of the BC [3]. IDC, also known as infiltrating carcinoma, is accounted for about 80% of all invasive BC [4]. The relevant molecular mechanism of IDC is not elucidated and thus the optimal management of IDC has become increasingly complex. Further investigation is urgent to identify effective approaches for the diagnosis and treatment of IDC [5].

Long noncoding RNA (lncRNA) are a class of RNA molecules longer than 200 nucleotides [6]. They regulate gene expression at the transcriptional, post-transcriptional, and epigenetic levels and also play an important role in pathological processes such as neurological disorders, diabetes, tumors [7,8] and more. Recent studies have clarified vital roles of several lncRNAs in BC pathobiology. Accordingly, lncRNAs could be a potential biomarker for BC prognosis, diagnosis, and therapeutic management [9]. Although cell-specific expression of lncRNAs and their role in regulating the expression of oncogenic and tumor suppressor genes have been

shown in recent investigations, their underlying molecular mechanisms in IDC remains undetermined [10–12]. In recent years, a number of lncRNAs including LOC100288637 and RP11-48B3 were found to be associated with abnormal estrogen receptor (ER) signaling pathway in BC, suggesting that lncRNAs can be applied as prognosis biomarker in patients with breast cancer [13,14]. As mentioned previously, the information on their roles in IDC remains unclear. Therefore, in the presented study, we assessed the expression levels of LOC100288637 and RP11-48B3 in patients with IDC as well as BC cell lines and analyzed their correlation with clinicopathological parameters.

Materials and methods

Study population

A total of 51 invasive ductal carcinoma breast cancer patients who had received no chemotherapy or radiotherapy were included in this study. All subjects were selected from individuals referred to Faghihie Hospital affiliated to Shiraz University of Medical Sciences. Written informed consent was obtained from all patients, and the study was approved by the Ethics Committee of Shiraz University of Medical Sciences, Shiraz, Iran. The demographic, reproductive and clinical information as well as the pathological data of tumor samples were obtained from each participant in the current

study. Tab. 1 shows the characteristics of the study participants. The participants were 27–68 years old, 15 (30%) participants were aged < 40 years, and 36 (70%) were ≥ 40 years old. Among the 51 participants, 36 (70%) were parous and 15 (30%) were nulliparous. Regarding the menopausal status, 4 (8%) participants were premenopausal and 47 (92%) were in postmenopausal status. Moreover, the participants were divided into 2 subgroups based on the age at menarče of < 14 or ≥ 14 years. The age at the first full term pregnancy and breast-feeding duration were also recorded. Parous participants were divided into 2 subgroups according to the age at the first full-term pregnancy of < 25 or ≥ 25 years.

Breast tissue sampling

Fresh breast cancer tissues and their adjacent noncancerous tissues were snap-freeze in liquid nitrogen after resection and stored at -80°C until RNA extraction.

Estrogen receptor, progesterone receptor and HER2 determination

The estrogen receptor (ER), progesterone receptor (PR), and HER2 status were determined according to the patient's histopathological data, following immunohistochemistry (IHC) staining. If $\geq 1\%$ of tumor cells show positive ER/PR staining, the ER/PR interpretation is positive. The IHC HER2 score 3+ was considered positive as well.

Cell culture

For the purpose of this study, we chose MCF7 and MDA-MB-231 (triple negative) human breast cell lines. All cells were obtained from the cell bank of Pasteur Institute of Iran, and cultured in RPMI 1640 medium (Sigma 42 Aldrich, St. Louis, MO, USA), supplemented with 10% fetal bovine serum (Gibco, Carlsbad, CA, USA), 100 U/ml penicillin, and 100 µg/ml streptomycin at 37 °C in 5% CO₂ at 95% humidity.

RNA isolation and real time polymerase chain reaction

Total RNA was extracted from all tissue samples and cell lines using the TRizol as was recommended by the manufacturer (Life Technologies, Carlsbad, CA) and the final concentration was quantified using a NanoDrop at 260 and 280 nm. In our experiment, 500 ng RNA was used for each sample with a purity range 1.8–2 for both 260/280 and 260/230 nm.

The RNA integrity was confirmed by gel electrophoresis and to remove the probable DNA contamination, the total RNA was treated with DNase (Takara Bio Inc, Otsu, Japan) according to the manufacturer's instruction. The cDNA synthesis was carried out using approximately 500 ng of total RNA with the Prime Script-RT kit (Takara, Japan) according to the manufacturer's protocol. In the next step, real time polymerase chain reaction (PCR) was carried out in a QuantStudio™ 3 system (Applied Biosystems, USA by Thermo Fisher Scientific) with SYBR Premix Ex Taq II kit (Takara, Japan) according to the manufacturer's protocol. Real-time specific primer pairs used were as follows: RP11-48B3, forward 5'-CAAGCCCTGATCAACTAGGAATA-3', reverse 5'-GGAAAGTTGGTTGCTGTGTAAG-3'; LOC100288637, forward 5'-CTAAGCCCTGCTTCTGGTATG-3', reverse 5'-GGAGGCAGATCCAGTTCATTAG-3'; B2M, forward 5'-AGATGAGTATGCCTGCCGTG-3', reverse 5'-GCGGCATCTTCAAACCTCCA-3'. For each reaction, 10 µl SYBR® Green Master Mix with 0.8 µl (40 nM) Primer 1 and 0.8 µl (40 nM) Primer 2 with 2 µl was used and the final volume was adjusted to a total of 18 µl using distilled H₂O. All the reactions were carried

Tab. 1. Demographic and reproductive characteristics of the participant subjects.

Variables	Subgroup	Number	Valid percent
age (year)	< 40	15	29.4
	≥ 40	36	70.6
family history for cancer	positive	26	51
	negative	25	49
marital status	married	43	84.3
	single	8	15.7
parity status	parous	36	70.6
	nulliparous	15	29.4
age at first full term pregnancy (years)	< 25	38	74.5
	≥ 25	10	19.6
abortion history	positive	12	23.5
	negative	37	72.5
breastfeeding experience	positive	37	72.5
	negative	14	27.5
breastfeeding (months)	0–6	20	39.2
	6–24	20	39.2
	≥ 24	11	21.6
age at menarche (year)	< 14	36	70.6
	≥ 14	15	29.4
menstrual cycles	regular	43	84.3
	irregular	8	15.7
menopausal status	pre	4	7.8
	post	47	92.2
oral contraceptive pills consumption	positive	10	19.6
	negative	41	80.4

out in triplicates. The real time PCR was performed in the following conditions: at 95 °C for 30 s followed by 40 repetitive cycles at 95 °C for 30 s and then at 60 °C for 30 s. No template controls were included in each run. Relative mRNA expression levels of LOC100288637 and RP11-48B3 were normalized to β2-microglobulin expression level as a house-keeping gene. The expression level (i.e. fold change) for each gene was calculated using the 2^{-ΔΔCT} method.

Bioinformatic analysis

In the current study, we also conducted different bioinformatics analyses, mainly by using the data of the Cancer Ge-

nome Atlas (TCGA) [15], to get more information about RP11-48B3.4 and LOC100288637. In this regard, we investigated the expression correlation between these two lncRNAs and mRNAs in the TCGA-BRCA dataset through using TANRIC web server [16]. Subsequently, we used the possible correlated mRNAs for any possible interactions between these mRNAs and lncRNA using the lncRRI search web [17].

Statistical analysis

The data are expressed as the mean ± SD and analyzed with SPSS version 20.0 software. P < 0.05 are considered statistically significant. The com-

Tab. 2. Pathological data of the evaluated tumor samples from breast cancer patients.

Clinical characteristics	Subgroup	Number	Valid percent
tumor size (cm)	< 2 cm	17	33.3
	2–5 cm	33	64.7
	> 5 cm	1	2
estrogen receptor	positive	45	88.2
	negative	6	11.8
progesterone receptor	positive	34	66.7
	negative	17	33.3
HER-2/neu status	positive	21	41.2
	negative	30	58.8
histological grade	1	11	21.6
	2	25	49
	3	15	29.4
lymph node metastasis	involved	30	58.8
	free	21	41.2

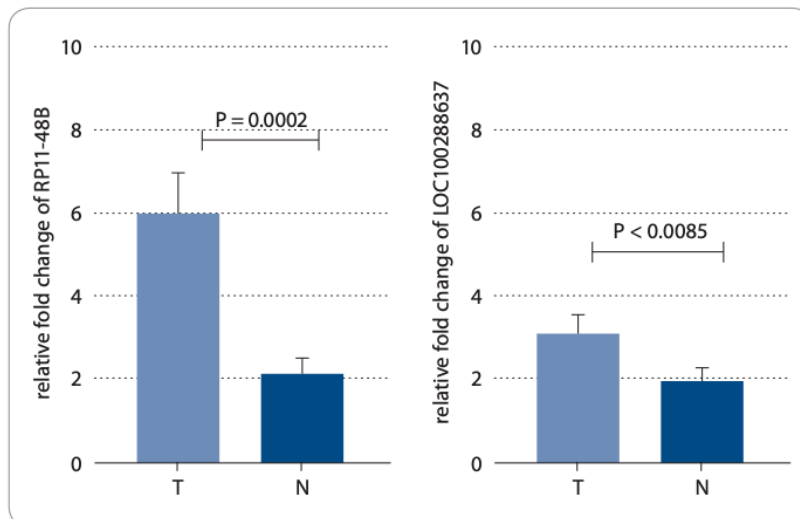


Fig. 1. Expression levels of lncRNAs LOC100288637 and RP11-48B3 in invasive ductal carcinomas and adjacent normal tissues. T and N denote tumor tissue and normal breast samples, respectively.

parisons between groups were made using unpaired Student's t-test or Mann-Whitney test. One-way ANOVA followed by Bonferroni or Kruskal-Wallis tests were used for multiple comparisons. The correlations were analyzed using Spearman's rank correlation coefficient.

Results

Demographics, reproductive and clinical data of the patients

The average patient age was 45.7 with the range 27–68 years. Fifty-one percent of the subjects had a family history of BC and other cancers while 49 % were negative for this variable. Among the 51 en-

rolled women, 36 participants were parous and 15 cases were nulliparous. The 37 women had experienced breastfeeding in their life and the rest were negative for the breastfeeding experience. Besides, 47 patients were in postmenopausal status and 4 patients were in premenopausal condition. Other demographics and reproductive characteristics of the patients are shown in Tab. 1. Clinically, all 51 tumor samples from BC patients had invasive ductal breast carcinoma. Most tumors were positive for ER (+45 vs. –6) and PR (+34 vs. –17) receptors, but negative for HER-2/neu marker (–30 vs. +21). A total of 33.3% of the tumor samples had a size < 2 cm, 64.7% had 2–5 cm, and 2% had > 5 cm. The tumor size was measured as the largest dimension of the microscopic invasive component in pathologic sections. The data on other tumor features including histological grade and lymph node metastasis (TNM) are indicated in Tab. 2. The tumor grade and TNM stage of the cells were determined by the WHO standard [18].

Expression levels of LOC100288637 and RP11-48B3 in BC tissues and different subgroups of the samples regarding the clinicopathological, demographic, and reproductive characteristics

In the next step, we investigated the expression profile of these two lncRNAs in 51 BC tissues and their adjacent normal tissues. The quantitative PCR (qPCR) results showed that LOC100288637 ($P = 0.0085$) was significantly upregulated in tumor tissues compared to normal samples. Furthermore, RP11-48B3 had a similar significant overexpression pattern in tumor tissues versus normal tissues ($P = 0.0002$). Fig. 1 represents the results on relative expression levels of both LOC100288637 and RP11-48B3 in tumor tissues compared to normal samples. In the next step, we compared the expression levels of the two lncRNAs in different subgroups of the patients with the various demographic, reproductive, and clinicopathological features of the subjects (Tab. 3 and 4). The findings showed that the lncRNA RP11-48B3.4 is upregulated in pa-

Tab. 3. Quantitative expression data of target lncRNAs in tumor tissues in relation to demographic and reproductive variables.

Clinicopathological characteristics	Subgroups	LncRNA RP11-48B3.4			LncRNA LOC100288637		
		Mean	Std. error	P-value	Mean	Std. error	P-value
age (years)	< 40	6.255	8.430	0.709	2.428	3.129	0.535
	≥ 40	5.850	6.581		3.302	3.908	
family history for cancer	positive	4.946	7.014	0.169	2.804	4.168	0.169
	negative	7.033	7.151		3.296	3.177	
marital status	married	6.336	7.458	0.468	3.196	3.895	0.698
	single	3.9973	4.464		2.235	2.272	
age at first marriage (years)	< 18	4.041	6.103	0.133	3.074	2.958	0.641
	18–24	8.458	8.931		2.543	3.431	
	≥ 24	5.336	4.905		4.931	5.703	
parity	parous	6.915	7.731	0.137	3.1701	3.736	0.694
	nulliparous	3.698	4.713		2.746	3.676	
age at first full term pregnancy (years)	< 25	5.944	7.1748	0.577	2.884	3.106	0.761
	≥ 25	7.201	7.729		4.436	5.599	
abortion history	positive	6.155	5.836	0.709	5.019	5.522	0.161
	negative	3.076	7.653		2.543	2.764	
breastfeeding experience	positive	6.747	7.692	0.213	3.097	3.710	0.883
	negative	3.913	4.814		2.908	3.759	
breastfeeding (months)	0–6	4.562	5.249	0.520	2.551	3.392	0.479
	6–24	6.511	7.155		3.597	4.184	
	≥ 24	7.541	9.714		2.939	3.424	
age at menarche (years)	< 14	9.014	9.597	0.041	3.254	3.490	0.079
	≥ 14	5.971	10.597		2.543	4.209	
menstrual cycles	regular	5.197	6.295	0.223	3.027	3.899	0.437
	irregular	10.116	9.894		3.142	2.418	
menopausal status	pre	4.682	5.997	0.551	4.466	7.786	0.483
	post	6.078	7.219		2.924	3.260	
oral contraceptive pills consumption	positive	6.220	4.416	0.209	4.591	5.728	0.962
	negative	5.908	7.643		2.668	2.977	

tients with age at menarche < 14 years compared to patients with age at menarche ≥ 14 (P = 0.041). The expression of this lncRNA was not significantly different among the different subgroups of the demographic and reproductive variables of the participants. Also, it was not observed any significant data on the difference in expression of the lncRNA LOC100288637 among various levels of the demographic and reproductive characteristics. From the clinicopathological point of view, our analyses indi-

cated that lncRNA RP11-48B3.4 was significantly upregulated in tumors with a lower histological grade (grade 1) compared to tumor samples with higher grades, including grades 2 and 3 (P = 0.047). Regarding LOC100288637, its expression showed a higher level in HER-2 positive tumors than HER-2 negative tumors (P = 0.031). For other different subgroups of the clinical characteristics, any noteworthy results were not found among them in terms of differences in expression of the two studied

lncRNAs (Tab. 4). In another of our statistic evaluations, we determined the correlation of LOC100288637 and RP11-48B3 with clinicopathological features of the BC patients. The Spearman's correlation analysis disclosed that the expression level of RP11-48B3 was negatively correlated with histological grade (r = -282, P = 0.045). The Spearman's analysis did not find any other significant correlation between the expression level of these lncRNAs and the other studied variables (Tab. 5).

Tab. 4. Expression levels of the studied lncRNAs in the tumor samples of the patients regarding to the clinicopathological characteristics.

Clinicopathological characteristics	Subgroups	LncRNA RP11-48B3.4			LncRNA LOC100288637		
		Mean	Std. error	P-value	Mean	Std. error	P-value
tumor size (cm)	< 2 cm	6.966	8.262	0.169	2.654	3.583	0.639
	2–5 cm	5.612	6.549		3.207	3.828	
	> 5 cm	0.788	0.0		4.349	0.0	
estrogen receptor	positive	5.873	7.411	0.200	3.034	3.625	0.787
	negative	6.686	4.301		3.131	4.500	
progesterone receptor	positive	6.622	7.898	0.549	2.925	3.295	0.905
	negative	4.664	5.0666		3.286	4.471	
HER-2	positive	5.550	7.5420	0.579	3.411	3.897	0.031
	negative	6.262	6.870		1.546	2.212	
histological grade	1	9.647	6.698	0.047	4.042	3.633	0.077
	2	5.497	7.985		2.477	3.937	
	3	4.058	4.816		3.260	3.336	
lymph node metastasis	involved	5.323	6.428	0.438	2.388	2.980	0.079
	free	7.116	8.023		4.049	4.434	

Tab. 5. Correlation of lncRNAs LOC100288637 and RP11-48B3.4 expressions with pathological variables.

Variables	LncRNA RP11-48B3.4		LncRNA LOC100288637	
	r	P-value	r	P-value
age (y)	-0.161	0.260	-0.034	0.815
menopausal status	0.084	0.557	0.099	0.489
family history of breast cancer	-0.195	0.171	-0.195	0.171
tumor size (cm)	-0.115	0.422	0.123	0.391
histological grade	-0.282	0.045*	-0.056	0.696
lymph node metastasis	0.127	0.374	0.252	0.184
estrogen receptor	-0.190	0.181	-0.029	0.840
progesterone receptor	0.085	0.554	-0.017	0.906
HER2	-0.078	0.584	-0.043	0.763

* The correlation is significant at the level of 0.05 (two-tailed).

Expression levels of LOC100288637 and RP11-48B3 in MDA-MB-231 and MCF-7 cell lines

For getting more information about involving lncRNAs LOC100288637 and RP11-48B3 in the pathogenesis of IDC, especially metastasis, their expression

levels were compared in human BC cell lines MDA-MB-231 and MCF-7 BC. The quantitative PCR (qPCR) data showed that the two lncRNAs were overexpressed in both MDA-MB-231 and MCF-7 BC cell lines; nevertheless, both lncRNAs showed a higher expression pattern in MDA-MB-231 than MCF7 cell

line (P = 0.0013 and P = 0.0003, respectively) (Fig. 2).

Bioinformatic evaluations

Through expression correlation analysis, we found some correlations between RP11-48B3.4 and mRNA expression in TCGA-BRCA genes, with negative correlation for mRNA MR1 (r = -0.409, P = 0) and positive correlations for mRNAs MRPS28 (r = 0.426, P = 0), NSMCE2 (r = 0.444, P = 0), POLR2K (r = 0.436, P = 0), TCEB1 (r = 0.476, P < 10⁻⁴⁷), UQCRB (r = 0.418, P = 0), YWHAZ (r = 0.464, P = 0), ZBTB10 (r = 0.567, P = 0), and ZNF706 (r = 0.471, P = 0) (Fig. 3). Regarding the interaction between the correlated mRNAs and RP11-48B3.4, we found that only zinc finger and BTB domain containing 10 (ZBTB10) were predicted to have interaction with this lncRNA (Fig. 4). It was not any observable interaction between lncRNA LOC100288637 and evaluated mRNAs.

Discussion

Invasive ductal carcinoma (IDC), as the most frequent form of BC in women, has a high clinical burden due to its high invasive properties. Although recent in-

vestigations have made significant advances in revealing some key molecular mechanisms regarding its pathogenesis, it shows a very heterogeneous and complex etiology with many unknown aspects [19,20]. Thus, it is extremely expedient to clarify the underlying molecular mechanisms through which IDC develops. lncRNA transcripts are emerging as key players in cancer initiation and pathobiology of BC, with both oncogenic and tumor-suppressive roles. In this regard, new experimental studies have revealed some novel molecular mechanisms by which lncRNAs involved in BC malignancy, providing a new avenue of investigation for characterizing the different hallmarks of BC [21].

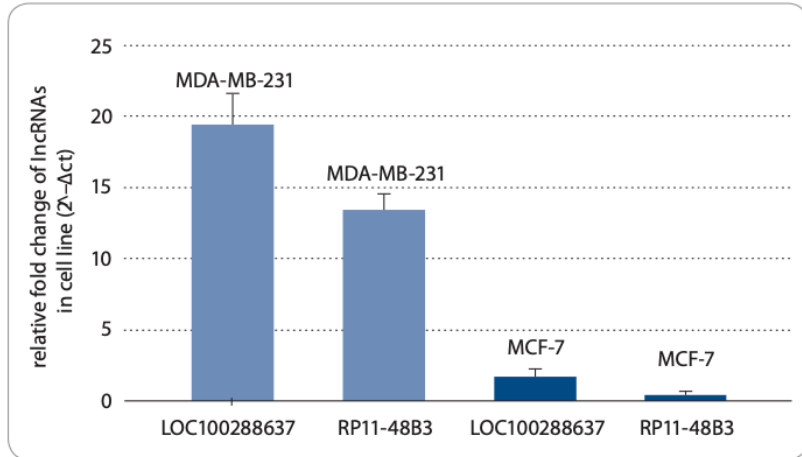


Fig. 2. Relative expression of target lncRNAs in MDA-MB-231 and MCF-7 cell lines compared to normal cell lines.

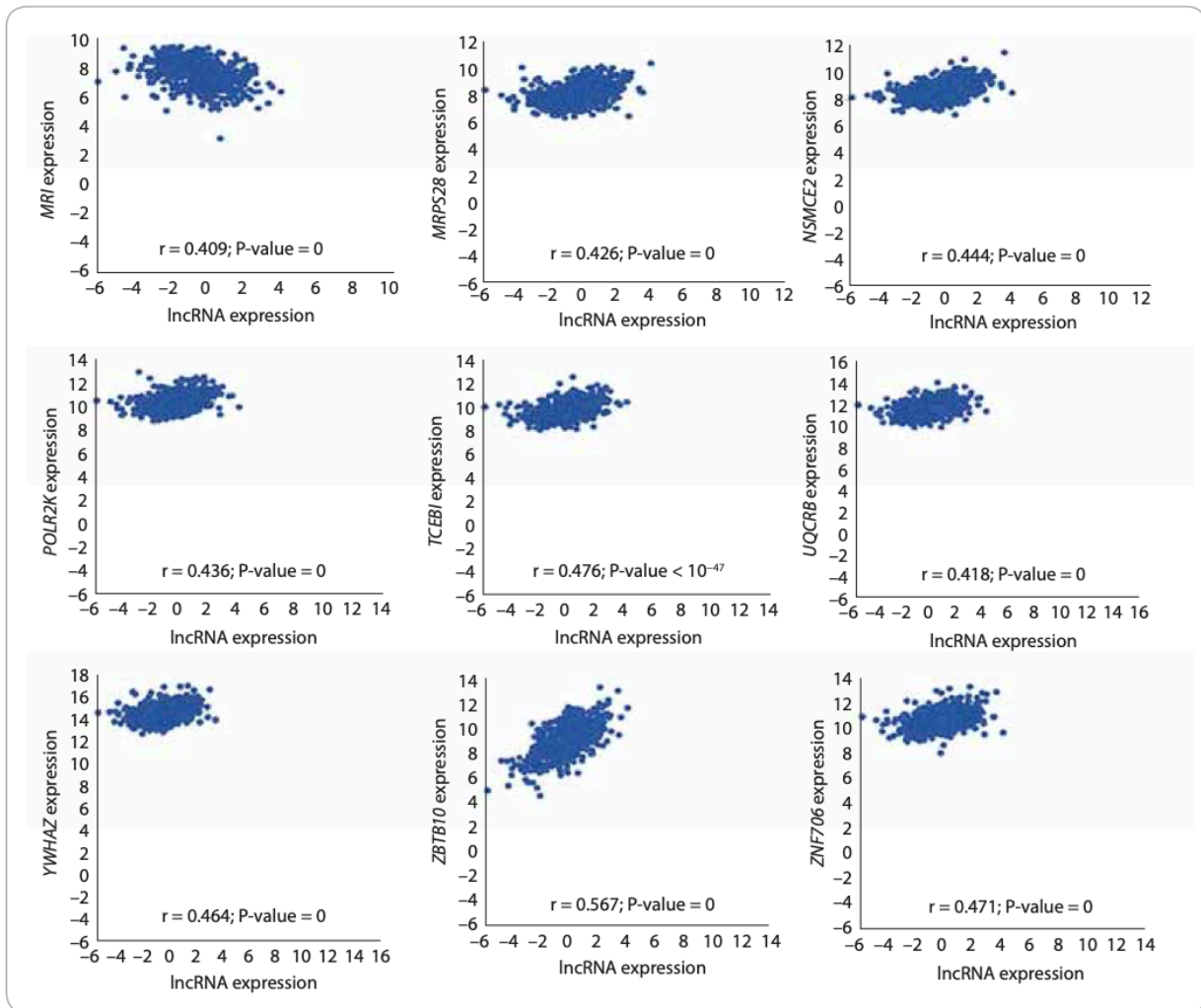


Fig. 3. Correlation analysis between RP11-48B3.4 and mRNAs expression in TCGA-BRCA.

288637 in HER-2 positive tumors compared to HER-2 negative samples. This finding is consistent with the results of previous work by Fan and colleagues that indicated the elevated expression level of LOC100288637 was strongly correlated with Her2/neu positive status in BC through next-generation sequencing and bioinformatics. Accordingly, we supposed that this lncRNA may play crucial roles in the pathogenesis of IDC via oncogenic functions. However, its exact mechanism needs more investigation by future studies. Furthermore, we observed a higher expression of the RP11-48B3 in lower grade tumors in comparison with the higher-grade tumors which also correlatively indicated by correlation analysis. Although this may be inconsistent with the abovementioned observation that this lncRNA had a higher expression in MDA-MB-231 cell line with high aggressive, invasive, and poorly differentiated properties, such a finding may occur due to our relatively small sample size or unknown complex nature of RP11-48B3 function during BC tumorigenesis which requires more investigation to disclose.

Interestingly, another result of the current study was that the lncRNA RP11-48B3.4 had an elevated expression pattern in BC patients with the age at menarche < 14 years in comparison to patients with the age at menarche ≥ 14 years. It has been reported that lower age at menarche increases the risk of BC through estrogen-related mechanisms [25,26]; however, little is known about its molecular mechanisms behind the risk of BC. Therefore, it can be suggested that a lower age at menarche may increase the risk of IDC partly through affecting the expression level of estrogen-linked lncRNA RP11-48B3.4. Although, confirmation of such assumptions requires conducting more functional studies.

Lastly, the current study also bioinformatically demonstrated some correlations between lncRNA RP11-48B3.4 and expression of several mRNAs in TCGA-BRCA, including negative expression correlation for mRNA MR1 and positive correlations for mRNAs *MRPS28*, *NSMCE2*, *POLR2K*, *TCEB1*, *UQCRB*, *YWHAZ*, *ZBTB10*,

and *ZNF706*. However, only the mRNA ZBTB10 were predicted to have interaction with the lncRNA RP11-48B3.4. Regarding LOC100288637, the results were not detected any evident interaction with given mRNAs. This highlights that the novel lncRNA RP11-48B3.4 may be involved in the pathogenesis of the IDC via interacting with some encoding genes by different mechanisms. There is evidence that lncRNAs play key roles in the development of BC through influencing the expression of other coding and non-coding genes [27]. In this way, one suggested mechanism is competing endogenous RNA (ceRNA) function through which lncRNAs regulate expression of mRNAs via sponging miRNAs in regulatory molecular networks whose roles in cancer development, especially in BC, are emerging with clinical significance [28,29].

Conclusion

The presented study showed the elevated expression levels of LOC100288637 and RP11-48B3 lncRNA in IDC breast tumors as well as BC cell lines have some important significance on their clinical outcome. This suggested them as putative oncogenic markers at the molecular level of IDC. However, it is important to analyze the correlation between their expression as well as progression-free survival time to conclude these lncRNAs as prognostic biomarkers. In our study, it was not possible to do such evaluation, since we used samples from newly diagnosed cancers. These findings can also be useful to candidate these two lncRNAs as targets for BC treatment. Future studies are warranted to analyze the expression of these two lncRNAs in various type of cancers to propose them as candidates of tumor biomarkers in combination with biomarker panels.

Acknowledgment

This study has been supported by the funding from the Shiraz University of Medical Sciences (No.16897). The authors would like to thank the staff members of Faghihi Hospital for their contributions to breast tissue sampling and data collection, and Autoimmune Disease Research Center of Shiraz University of Medical Science, for their help and support and also the participants who took part in this study.

Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Author contributions

A.D, Y.M, and S.I contributed to the conception and design of the research. F.F, S.I, M.D.F, and M.R.M performed the experiments. F.F, S.I, H.D, M.D. F, and Y. M interpreted the results of the experiments, analyzed data, and prepared the figures. S.I, M.R.M, and A.D drafted the manuscript. A.D, H.D, S.M.B.T, and S.I. edited and revised the manuscript. M.B.T and Y.M, made study supervision and technical or material support. All authors read and approved the final manuscript.

References

1. Bray F, Ferlay J, Soerjomataram I et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018; 68(6): 394–424. doi: 10.3322/caac.21492.
2. Cancer Genom Atlas Network. Comprehensive molecular portraits of human breast tumors. *Nature* 2012; 490(7418): 61–70. doi: 10.1038/nature11412.
3. Tulinius H, Bjarnason O, Sigvaldason H et al. Tumours in Iceland 10. Malignant tumors of the female breast: a histological classification, laterality, survival, and epidemiological considerations. *APMIS* 1988; 96(3): 229–238. doi: 10.1111/j.1699-0463.1988.tb05296.x.
4. Tan PH and Ellis IO. Myoepithelial and epithelial-myoepithelial, mesenchymal, and fibroepithelial breast lesions: updates from the WHO classification of tumours of the breast 2012. *J Clin Pathol* 2013; 66(6): 465–470. doi: 10.1136/jclinpath-2012-201078.
5. DeSantis CE, Lin CC, Mariotto AB et al. Cancer treatment and survivorship statistics, 2014. *CA Cancer J Clin* 2014; 64(4): 252–271. doi: 10.3322/caac.21235.
6. Carninci P, Kasukawa T, Katayama S et al. The transcriptional landscape of the mammalian genome. *Science* 2005; 309(5740): 1559–1563. doi: 10.1126/science.1112014.
7. Du T, Zhang B, Zhang S et al. Decreased expression of long non-coding RNA WT1-AS promotes cell proliferation and invasion in gastric cancer. *Biochim Biophys Acta* 2016. 1862(1): 12–19. doi: 10.1016/j.bbdis.2015.10.001.
8. Kurokawa R. Long noncoding RNAs: structures and functions. Springer 2015.
9. Iranpour M, Soudyab M, Geranpayeh L et al. Expression analysis of four long noncoding RNAs in breast cancer. *Tumour Biol* 2016; 37(3): 2933–2940. doi: 10.1007/s13277-015-4135-2.
10. Nie Z-L, Wang Y-S, Mei Y-P et al. Prognostic significance of long noncoding RNA Z38 as a candidate biomarker in breast cancer. *J Clin Lab Anal* 2018; 32(1): e22193. doi: 10.1002/jcla.22193.
11. Calin GA, Liu C-G, Ferracin M et al. Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell* 2007; 12(3): 215–229. doi: 10.1016/j.ccr.2007.07.027.
12. Gupta RA, Shah N, Wang KC et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 2010; 464(7291): 1071–1076. doi: 10.1038/nature08975.
13. Wu L, Xu Q, Zhang H et al. A new avenue for obtaining insight into the functional characteristics of long noncoding RNAs associated with estrogen receptor signaling. *Sci Rep* 2016; 6: 31716. doi: 10.1038/srep31716.
14. Yang F, Lyu S, Dong S et al. Expression profile analysis of long noncoding RNA in HER-2-enriched subtype breast cancer by next-generation sequencing and

SUPPLEMENTAL INFORMATION

Table S1. Cancer drivers with proto-oncogenic and tumor suppressor roles in cluster 8

GeneID	Oncogene	Tumor Suppressor	epi_flag	TF_flag	logFC. D2	logFC. D7	logFC. D30	logFC. D46
POU5F1	1	0	not_epi	TF	-2,398	-10,062	-12,15	-9,700
USP44	1	0	epiFactor	not_TF	-3,731	-5,234	-8,680	-8,395
LCK	1	0	not_epi	not_TF	-5,825	-12,451	-9,767	-6,898
KDR	1	0	not_epi	not_TF	-2,099	-1,811	-1,034	-3,144
KAT7	1	0	epiFactor	not_TF	-0,981	-1,384	-1,959	-1,997
BCOR	1	0	epiFactor	not_TF	-0,863	-2,012	-1,542	-1,957
TERT	1	0	not_epi	not_TF	-3,219	-4,013	-6,944	-10,51
MAP2K1	1	0	not_epi	not_TF	-0,897	-1,483	-1,687	-1,211
VAV1	1	0	not_epi	not_TF	-6,920	-10,113	-10,11	-4,636
RET	1	0	not_epi	not_TF	-1,886	-3,810	-2,790	-3,140
ETV4	1	0	not_epi	TF	-0,171	-5,140	-3,275	-3,035
MYC	1	0	not_epi	TF	-2,261	-1,763	-3,087	-2,138
CARD11	1	0	not_epi	not_TF	-1,518	-2,832	-3,582	-2,451
FGFR4	1	0	not_epi	not_TF	-1,001	-2,831	-2,707	-2,721
ERCC2	1	0	not_epi	not_TF	-0,907	-0,999	-1,588	-1,997
AXIN1	1	1	not_epi	not_TF	-0,197	-1,015	-1,406	-1,225
NTHL1	1	0	not_epi	not_TF	-1,200	-1,543	-2,053	-1,921
ARHGAP26	1	0	not_epi	not_TF	-0,497	-1,248	-1,108	-0,925
PSIP1	1	0	epiFactor	not_TF	-1,215	-0,762	-2,119	-1,704
MYCL	1	0	not_epi	not_TF	-0,977	-1,693	-2,523	-3,187
OLIG2	1	0	not_epi	TF	-1,165	-1,892	-2,856	-4,042
TEC	1	0	not_epi	not_TF	-2,254	-4,295	-3,257	-2,423
GRM3	1	0	not_epi	not_TF	-5,432	-4,193	-2,847	-4,241
PTPN6	1	1	not_epi	not_TF	-0,767	-2,015	-2,381	-1,817
FOXL2	1	1	not_epi	TF	-1,212	-3,767	-7,811	-4,694
CTNND2	1	0	not_epi	not_TF	-0,832	-0,312	0,675	-2,265
RGS7	1	0	not_epi	TF	-0,557	-2,834	-0,426	-3,632
BCL11A	1	0	not_epi	TF	-0,405	-2,595	-0,650	-1,715
WNK2	1	1	not_epi	not_TF	-2,300	-3,944	-2,020	-2,641

Table S2. Involvement of several TF genes as proto-oncogenes extracted from clusters 4 and 8 with altered expression in most of muscle differentiation stages.

GeneID	logFC.D2	logFC.D7	logFC.D30	logFC.D46
POU5F1	-2,40	-10,06	-12,15	-9,70
ZNF331	-0,36	-0,59	-2,40	-2,38
MYCN	-1,86	-0,32	-2,52	-5,31
HMGA1	-0,64	-1,24	-4,68	-4,51
ETV4	-0,17	-5,14	-3,27	-3,03
MYC	-2,26	-1,76	-3,09	-2,14
PATZ1	-0,19	-0,27	-0,99	-1,23
OLIG2	-1,17	-1,89	-2,86	-4,04
RGS7	-0,56	-2,83	-0,43	-3,63
MYB	-0,91	0,23	-2,03	-2,30
BCL11A	-0,40	-2,60	-0,65	-1,71

Table S3. TF genes as only proto-oncogenic role from clusters 4, 5, 7, and 8 across muscle differentiation stages.

GeneID	logFC.timeD2	logFC.timeD7	logFC.timeD30	logFC.timeD46
POU5F1	-2,40	-10,06	-12,15	-9,70
LEF1	11,67	10,60	7,40	7,32
ZNF331	-0,36	-0,59	-2,40	-2,38
MYCN	-1,86	-0,32	-2,52	-5,31
HMGA1	-0,64	-1,24	-4,68	-4,51
ETV4	-0,17	-5,14	-3,27	-3,03
PAX3	7,71	11,25	7,74	6,59
MYC	-2,26	-1,76	-3,09	-2,14
PATZ1	-0,19	-0,27	-0,99	-1,23
OLIG2	-1,17	-1,89	-2,86	-4,04
RGS7	-0,56	-2,83	-0,43	-3,63
MYB	-0,91	0,23	-2,03	-2,30
HMGA2	0,08	2,44	1,27	-0,43
BCL11B	1,75	-1,55	0,86	-0,54
CREB3L2	1,10	1,68	1,38	0,94
DDIT3	0,53	-1,24	-0,33	0,08
BCL11A	-0,40	-2,60	-0,65	-1,71

Table S4. Genes with Epifactor role with proto-oncogenic function from in clusters 4, 7, and 8

GeneID	logFC.timeD2	logFC.timeD7	logFC.timeD30	logFC.timeD46
USP44	-3,73	-5,23	-8,68	-8,39
SMARCD1	-0,12	-0,31	-1,50	-2,04
KAT7	-0,98	-1,38	-1,96	-2,00
MSH6	-0,15	-0,27	-1,57	-1,59
BCOR	-0,86	-2,01	-1,54	-1,96
SET	-0,07	-0,54	-2,14	-2,17
YWHAE	0,09	0,02	-1,14	-1,29
DAXX	0,32	-0,07	-1,21	-1,05
SRSF3	0,24	0,01	-1,09	-1,16
TET1	-0,60	-1,33	-2,09	-2,34
BAZ1A	2,00	1,85	0,78	0,83
PSIP1	-1,21	-0,76	-2,12	-1,70

Table S5. Genes without TF and Epifactor roles which function as proto-oncogenes.

GeneID	cluster	logFC.timeD2	logFC.timeD7	logFC.timeD30	logFC.timeD46
LCK	8	-5,82	-12,45	-9,77	-6,90
BUB1B	4	-0,97	-0,63	-2,89	-2,65
RMI2	4	-1,48	-1,02	-3,04	-2,37
RHOH	4	1,36	-4,35	-7,29	-6,88
COL2A1	7	1,79	5,55	2,78	1,47
KDR	8	-2,10	-1,81	-1,03	-3,14
FANCD2	4	-0,25	-0,21	-2,39	-1,99
KCNJ5	5	8,38	4,62	2,54	1,45
RFWD3	4	0,27	-0,24	-1,88	-1,87
IRS4	4	-2,37	-1,85	-5,09	-5,58
FEN1	4	0,83	0,09	-2,00	-1,82
LMO2	5	6,09	1,47	1,74	1,83
BCL2L12	4	0,24	-0,92	-2,20	-2,03
FANCA	4	0,36	-0,56	-1,89	-1,30
CCNE1	4	0,34	-0,48	-1,75	-1,73
HIP1	4	-0,06	0,02	-1,29	-1,30
TERT	8	-3,22	-4,01	-6,94	-10,51
SMC1A	4	0,30	0,03	-1,34	-0,82
MAP2K1	8	-0,90	-1,48	-1,69	-1,21
CXCR4	5	6,80	2,91	2,64	-0,04
RNF43	5	4,12	0,89	0,12	0,39

VAV1	8	-6,92	-10,11	-10,11	-4,64
PPM1D	7	0,00	1,54	0,04	-0,33
SRSF2	4	0,21	0,03	-1,06	-1,42
CCNB1IP1	4	0,28	-0,89	-1,66	-1,86
MTOR	4	0,09	-0,24	-1,36	-1,45
TAF15	4	0,09	-0,13	-1,21	-1,20
AFDN	4	0,53	0,32	-0,97	-1,84
DDX10	4	0,79	0,37	-1,05	-0,98
FUBP1	4	-0,05	0,13	-1,05	-1,39
PABPC1	4	0,28	-0,59	-1,33	-1,03
HNRNPA2B1	4	-0,16	-0,05	-1,29	-1,24
RECQL4	4	0,47	-0,38	-1,88	-1,09
PCBP1	4	-0,18	-0,80	-1,37	-1,16
KNSTRN	4	-0,03	0,29	-1,01	-1,33
RET	8	-1,89	-3,81	-2,79	-3,14
STIL	4	0,76	0,81	-1,02	-1,02
CNBP	4	0,23	-0,32	-1,30	-1,12
CARD11	8	-1,52	-2,83	-3,58	-2,45
FGFR4	8	-1,00	-2,83	-2,71	-2,72
MSI2	7	0,32	0,68	-0,04	-1,07
ERCC2	8	-0,91	-1,00	-1,59	-2,00
FLT4	4	-1,61	-0,65	-2,27	-2,82
NTHL1	8	-1,20	-1,54	-2,05	-1,92
LEPROTL1	4	0,64	0,65	-0,47	-1,36
POLD1	4	-0,18	-0,56	-2,18	-1,49
POLQ	4	0,40	0,60	-1,34	-0,78
ARHGAP26	8	-0,50	-1,25	-1,11	-0,93
ATP1A1	4	0,49	-0,12	-1,05	-0,79
MYCL	8	-0,98	-1,69	-2,52	-3,19
TEC	8	-2,25	-4,29	-3,26	-2,42
GRM3	8	-5,43	-4,19	-2,85	-4,24
PIM1	4	-1,90	-0,11	-2,38	-2,07
TPM3	4	0,26	-0,45	-1,29	-1,18
PREX2	4	-1,17	-0,68	-2,25	-1,84
PIK3CB	4	-0,75	-0,24	-1,74	-1,30
CTNND2	8	-0,83	-0,31	0,68	-2,26

Table S6. Tumor suppressor genes without TF and Epifactor roles.

GeneID	cluster	logFC.timeD2	logFC.timeD7	logFC.timeD30	logFC.timeD46
EMP1	1	-2,88	3,62	7,10	8,32
SRPX	9	0,25	0,39	2,80	4,78
GAS1	1	0,64	4,11	6,31	6,27
PCDH9	2	3,20	4,38	6,61	5,40
NUPR1	9	2,62	0,14	7,60	8,75
VIM	2	3,11	2,84	4,79	4,66
WNT5A	2	8,19	6,47	5,44	4,86
AHNAK	1	2,58	5,74	6,11	6,71
DMD	1	-0,23	3,40	3,45	3,22
IGFBP5	1	-1,30	5,07	7,05	7,26
SLIT2	2	2,49	4,35	3,30	2,79
DAB2	2	1,76	2,54	4,02	4,97
PLCE1	2	3,38	6,43	6,06	7,12
EPHB3	1	-0,56	3,77	5,29	5,73
SPARCL1	9	-3,90	1,13	6,80	7,62
INPP4B	1	-0,82	5,13	4,85	4,64
PDGFRL	9	-4,31	-1,00	2,72	4,38
ROR2	2	3,78	2,87	2,99	2,93
CLU	9	-1,40	-0,69	0,85	2,91
BLCAP	9	-1,21	-0,56	0,95	1,10
CXCL14	9	-0,43	2,04	3,30	6,14
RBMS3	1	0,94	3,45	4,83	5,24
PLD1	1	-3,04	1,56	3,88	3,46
AGTR1	9	0,73	2,08	7,86	9,24
GREM1	2	5,76	9,94	7,17	7,92
CADM1	1	-0,25	2,36	3,10	2,06
RECK	1	0,06	1,84	3,04	3,14
IGFBP4	9	0,50	-1,13	1,52	2,36
GBP1	9	-0,86	1,68	9,03	8,96
DLK1	1	1,11	9,91	12,92	13,87
PLA2R1	9	0,62	0,29	5,43	5,13
ZNF185	1	-1,68	1,53	1,25	0,78
GSN	9	0,13	0,99	2,58	3,32
LSAMP	1	-0,50	3,87	5,06	5,23
CDH13	1	1,25	5,21	5,44	5,41

CXCL12	9	-2,14	0,35	2,70	1,78
SAMD9L	9	0,00	3,12	9,44	11,55
RBL2	9	0,05	0,68	1,53	1,82
SYNPO2	9	-4,03	-0,97	2,98	2,58
CXXC4	1	-2,08	3,90	3,12	1,03
PDLIM4	2	1,48	1,36	3,16	3,19
ANGPTL4	9	-2,40	-0,18	4,19	4,64
NDN	9	-0,39	0,48	1,87	1,83
DLC1	2	4,64	4,61	4,96	4,94
RND3	1	-1,31	1,05	1,25	2,01
HPGD	9	-4,15	-2,98	1,32	2,66
DPP4	9	-2,95	-0,44	3,62	4,98
TOM1L2	1	-0,41	2,12	3,31	3,07
TCEAL7	1	-0,92	1,32	1,40	1,76
IGFBP3	1	-2,23	3,51	7,30	6,09
BMPR2	9	0,92	1,15	2,99	2,77
SOCS3	9	-2,32	-1,25	-0,02	2,25
ADAMTS9	2	2,90	5,44	6,54	6,71
UNC5C	2	6,38	5,91	6,07	5,69
AHRR	9	0,15	1,76	3,77	4,13
TXNIP	9	-0,65	-0,78	2,43	2,48
EFNA5	1	-0,22	1,64	2,44	2,38
WNT7A	9	1,88	4,59	5,65	9,36
TNFSF12	9	-1,98	-0,56	2,21	2,77
FBXO25	9	-0,50	0,29	1,42	1,52
CDKN1C	1	1,43	2,37	4,18	4,49
FHL1	9	-0,41	0,38	2,77	2,03
HTRA1	1	-1,43	2,59	5,84	4,82
SEMA3B	1	0,04	2,19	2,82	2,71
LOX	9	-4,83	0,77	6,95	6,06
TGFBI	1	0,28	3,60	7,01	5,72
EPB41L3	2	1,94	1,22	1,79	1,98
PCDHGC3	2	1,48	2,16	2,24	2,07
DCDC2	9	1,01	2,15	4,29	6,29
RNASEL	1	0,74	1,79	3,01	3,32
BTG2	2	3,23	1,54	1,89	2,91
HECA	9	-0,31	0,59	1,88	1,95

PCDH17	1	1,69	3,33	4,70	2,95
CLDN1	9	0,78	0,32	2,57	4,64
VEGFA	9	-0,89	-1,37	0,89	1,76
IGF1	1	0,00	8,01	12,93	12,44
ALDH1A2	2	11,64	7,59	10,63	11,23
SCUBE2	9	-1,03	2,71	4,81	5,95
XAF1	9	-3,93	-0,82	4,19	5,77
DCN	1	-1,95	11,43	15,02	15,70
THSD1	1	0,68	2,82	3,08	3,98
SPARC	1	0,26	1,81	3,55	3,37
CMTM3	2	1,46	2,08	2,00	2,11
SEC14L2	9	0,86	0,05	1,52	2,51
IGFBP7	1	1,49	5,05	9,47	8,37
TNFSF9	9	0,53	0,19	1,72	3,21
RASSF4	1	-2,43	1,41	2,59	2,50
MYO18B	9	0,58	-1,54	4,24	4,74
SPRY2	2	1,17	0,41	1,15	2,12
TIMP3	1	-1,29	3,18	4,89	4,55
TP53INP1	2	1,38	2,36	3,75	2,97
PRKCE	1	-0,61	2,02	3,52	3,68
SELENBP1	1	0,32	1,46	2,27	2,80
EPHB2	9	-1,93	-0,06	1,55	1,60
ADAMTS18	1	-0,76	7,92	5,04	4,01
DSC3	9	-1,33	0,54	3,14	1,77
TGFBR3	9	0,99	2,37	3,47	5,53
S100A2	9	-0,40	0,27	2,07	4,64
ANXA1	1	1,39	5,45	4,48	4,37
ARHGAP29	2	1,97	3,52	3,28	4,71
PDCD4	1	-0,46	1,07	1,40	1,59
CEBPD	9	-2,52	-0,64	3,05	3,91
ARL6IP5	1	-0,34	0,78	1,56	2,03
MT1M	9	3,18	0,00	3,21	8,11
HSPB7	9	-2,75	4,70	7,78	7,89
RHOBTB2	9	0,68	0,79	1,43	1,78
CD44	2	3,22	1,56	3,33	4,30
MRVI1	1	-0,67	0,43	1,73	0,58
HACD4	9	0,05	0,47	1,74	2,62

DIRAS3	2	3,50	4,98	5,30	5,34
NAPEPLD	1	0,39	1,81	2,44	2,56
MAL	9	0,00	0,00	8,37	11,07
CTNNA3	2	2,67	3,43	3,84	4,04
GPRC5A	9	-6,38	-0,25	3,72	6,60
MARVELD1	9	0,14	0,40	1,29	1,38
BNIP3L	9	-0,60	0,20	1,09	1,20
RNF144A	9	0,05	0,58	0,92	1,55
UBE2QL1	9	0,27	-1,02	2,91	0,76
PLK2	1	0,37	1,10	1,47	1,51
LIFR	2	1,83	0,66	3,53	3,06
NNAT	1	0,40	1,54	2,76	2,43
PIAS1	1	-0,16	0,83	0,94	1,01
TMEFF2	2	3,10	2,63	4,61	4,59
DKK3	9	-0,28	0,62	3,21	2,73
HEPACAM	9	1,88	2,61	4,78	7,65
ITGB3	9	-3,26	3,25	6,63	7,40
FRK	9	-0,32	-0,95	2,05	3,21
MARCKS	2	0,73	1,16	2,01	1,92
SYNM	1	-0,60	1,69	0,89	0,48
ST5	1	-0,61	0,93	1,51	1,30
YPEL3	9	-1,42	-0,84	0,67	0,89
NDRG2	9	-0,98	0,18	1,89	1,89
FBXO32	9	-2,15	0,75	2,56	3,81
PHLDA3	1	-1,54	0,74	2,04	2,18
MAP3K8	2	1,94	1,50	1,36	1,73
NOTCH3	1	-1,10	0,54	1,46	0,64
WNT11	2	5,20	3,74	6,27	7,14
CADM3	1	-0,77	1,26	1,81	0,45
TUSC1	2	1,72	1,45	1,99	2,09
SH3GLB1	2	0,83	1,14	1,92	1,66
ACHE	1	-1,10	0,96	2,00	1,55
BLNK	9	-0,05	-1,94	5,48	6,75
EMP2	2	0,95	2,09	1,09	2,35
ZBTB18	2	0,52	0,92	1,20	0,40
FBLN1	2	1,26	1,47	1,40	1,25
SASH1	9	0,25	0,41	1,01	1,20

GLIPR1	2	5,85	5,40	6,72	5,22
CAMK2N1	9	-1,26	-0,08	1,68	1,45
CFTR	1	-2,57	2,63	4,27	2,64
APAF1	2	0,87	1,74	1,56	1,53
EXTL1	9	0,51	0,98	3,67	3,91
MT2A	9	-1,17	-0,12	2,37	4,60
SSBP2	1	-0,14	2,14	2,42	2,56
GJB2	9	-2,03	0,11	5,56	5,77
EDNRB	1	-0,06	6,19	4,85	4,25
CAV1	9	0,22	0,36	2,24	2,45
SRGAP3	1	-3,22	1,94	2,98	2,72
WFDC1	9	0,29	-1,09	4,56	2,35
CD82	2	2,43	0,87	2,83	2,87
NEURL1	1	0,84	4,28	4,04	4,47
BIN1	9	0,09	0,22	1,09	1,66
THBD	1	1,14	4,37	5,50	5,33
PTCH2	2	2,47	1,73	3,41	3,79
ZMYND10	9	0,67	1,45	2,82	3,45
GORAB	1	0,08	0,46	0,97	1,07
PCDH10	2	2,68	1,64	4,21	1,24
IFT88	2	1,29	1,97	2,23	2,09
CDKN2B	1	3,82	7,27	10,18	6,85
MAPK10	1	-0,49	2,17	1,44	0,71
PPP3CC	2	0,82	1,30	1,66	1,43
AFAP1L2	2	1,94	2,36	2,76	3,07
SHISA3	9	-0,03	0,50	4,03	3,78
KLK6	9	-0,37	1,19	-0,05	2,93
PACRG	9	1,18	1,11	3,36	3,11

Table S7. Tumor suppressor genes and proto-oncogenes without TF and Epifactor roles among all clusters.

GeneID	cluster	Oncogene	Tumor Suppressor	logFC. D2	logFC. D7	logFC. D30	logFC. D46
EPHA3	1	1	1	-3,58	3,10	5,83	5,23
TGFBR2	9	1	1	-1,04	0,04	2,54	3,87
CDH11	2	1	1	7,30	5,81	6,93	6,93
FAT4	1	1	1	-0,13	3,93	6,32	6,96
CDKN2C	9	1	1	1,25	3,14	6,65	8,43
DCC	9	1	1	1,49	1,03	4,19	2,63
BTK	5	1	1	3,15	-1,60	-1,84	-6,34
FAS	9	1	1	0,63	2,22	3,22	5,15
WIF1	9	1	1	-2,32	-1,64	2,92	3,41
NOTCH2	1	1	1	-0,10	1,85	1,75	1,85
LRIG3	2	1	1	1,94	3,69	3,30	2,77
CHEK2	4	1	1	-0,55	-0,90	-2,09	-2,02
BLM	4	1	1	0,42	0,39	-2,04	-2,18
PTPRD	1	1	1	-0,23	2,22	1,96	0,97
AXIN2	5	1	1	3,94	1,83	0,21	-0,16
PTPRT	9	1	1	0,45	-1,74	3,58	2,41
CDKN1A	1	1	1	0,89	2,81	4,07	5,17
PALB2	4	1	1	0,14	-0,35	-1,29	-1,42
MSH2	4	1	1	-0,92	-1,24	-2,84	-2,43
DDR2	2	1	1	1,59	1,98	2,70	2,92
IDH1	4	1	1	0,18	-1,51	-2,15	-1,81
GPC5	1	1	1	-1,26	4,44	4,86	3,50
SFRP4	1	1	1	-1,37	8,07	3,57	5,73
SYK	6	1	1	0,78	-3,50	1,86	1,87
TNFAIP3	1	1	1	0,83	2,30	2,88	3,24
CDH1	6	1	1	-4,12	-6,93	-3,55	-1,65
AXIN1	8	1	1	-0,20	-1,02	-1,41	-1,22
LRP1B	1	1	1	1,90	8,19	10,54	10,37
SOCS1	6	1	1	-3,45	-2,60	-1,23	-0,87
PRKAR1A	2	1	1	0,49	0,73	1,15	1,26
NDRG1	9	1	1	0,98	0,57	2,25	3,04
PTCH1	7	1	1	0,72	1,45	-0,56	-0,71
CYLD	1	1	1	0,02	0,84	1,39	1,45
SUFU	2	1	1	1,31	0,89	1,41	1,56
GPC3	1	1	1	0,83	2,85	2,83	2,39
CASP8	2	1	1	0,92	1,06	1,35	1,83

NOTCH1	5	1	1	0,76	-1,06	-0,57	-1,19
CBL	4	1	1	-0,11	-0,38	-0,99	-1,30
PTPN6	8	1	1	-0,77	-2,01	-2,38	-1,82
PTPRC	9	1	1	-3,63	-1,80	3,77	2,14
CDK6	7	1	1	2,15	2,95	2,33	1,68
ITGAV	1	1	1	0,14	2,66	2,85	2,67
FANCG	4	1	1	-0,42	-0,20	-1,19	-0,69
NCOA4	2	1	1	1,02	0,80	0,85	0,88
FHIT	9	1	1	0,46	0,13	1,96	0,98
FAT1	1	1	1	-0,08	0,59	1,41	1,27
AMER1	5	1	1	1,29	0,24	-0,27	-0,12
CNOT3	4	1	1	-0,12	-0,46	-0,92	-1,06
SH2B3	4	1	1	-0,18	-1,50	-1,54	-2,02
EXT2	9	1	1	0,23	0,44	1,02	1,14
PML	9	1	1	-0,17	-0,38	0,91	1,06
CDKN2A	2	1	1	2,95	4,37	5,56	5,20
FLCN	9	1	1	0,70	-0,20	0,25	1,12
LATS2	3	1	1	-1,22	0,75	0,34	0,09
WNK2	8	1	1	-2,30	-3,94	-2,02	-2,64
EXT1	1	1	1	0,08	1,43	2,04	1,24

Table S8. GO:BP terms for the expression of genes with both TF and Epifactor roles across muscle differentiation stages.

Index	Name	P-value	Adjusted p-value
1	regulation of transcription, DNA-templated (GO:0006355)	2.778e-9	4.027e-7
2	negative regulation of vitamin D biosynthetic process (GO:0010957)	0.002248	0.02107
3	regulation of vitamin D receptor signaling pathway (GO:0070562)	0.002248	0.02107
4	regulation of cardiac muscle hypertrophy in response to stress (GO:1903242)	0.002697	0.02107
5	negative regulation of vitamin metabolic process (GO:0046137)	0.002697	0.02107
6	cellular response to cold (GO:0070417)	0.002697	0.02107
7	regulation of vitamin D biosynthetic process (GO:0060556)	0.002697	0.02107
8	desmosome organization (GO:0002934)	0.003146	0.02107
9	negative regulation of stress-activated protein kinase signaling cascade (GO:0070303)	0.003146	0.02107
10	brown fat cell differentiation (GO:0050873)	0.003595	0.02107

Table S9. GO:BP terms for genes with both TF and Epifactor roles with only altered expression in Day 2.

Index	Name	P-value	Adjusted p-value
1	negative regulation of transcription, DNA-templated (GO:0045892)	0.0001062	0.001805
2	regulation of transcription by RNA polymerase II (GO:0006357)	0.001340	0.005394
3	regulation of transcription, DNA-templated (GO:0006355)	0.001411	0.005394
4	histone H2A-K119 monoubiquitination (GO:0036353)	0.001200	0.005394
5	negative regulation of nucleic acid-templated transcription (GO:1903507)	0.001586	0.005394
6	histone H2A monoubiquitination (GO:0035518)	0.002548	0.006188
7	negative regulation of cellular macromolecule biosynthetic process (GO:2000113)	0.002199	0.006188
8	negative regulation of transcription by RNA polymerase II (GO:0000122)	0.003424	0.007277
9	negative regulation of G0 to G1 transition (GO:0070317)	0.005689	0.01043
10	regulation of G0 to G1 transition (GO:0070316)	0.006138	0.01043

Table S10. GO:BP terms for genes with both TF and Epifactor roles with altered expression at somite stage.

Index	Name	P-value	Adjusted p-value
1	positive regulation of myoblast proliferation (GO:2000288)	0.001249	0.01574
2	regulation of mitochondrial membrane permeability involved in apoptotic process (GO:1902108)	0.001749	0.01574
3	macromolecule modification (GO:0043412)	0.001999	0.01574
4	positive regulation of mitochondrial membrane permeability involved in apoptotic process (GO:1902110)	0.002248	0.01574
5	covalent chromatin modification (GO:0016569)	0.003495	0.01747
6	mitotic intra-S DNA damage checkpoint signaling (GO:0031573)	0.003745	0.01747
7	positive regulation by host of viral transcription (GO:0043923)	0.003994	0.01747
8	modulation by host of symbiont process (GO:0051851)	0.007975	0.03101
9	chromatin organization (GO:0006325)	0.0004936	0.01574
10	ATP-dependent chromatin remodeling (GO:0043044)	0.01046	0.03660

Table S11. GO:BP terms for genes with both TF and Epifactor roles with their altered expression at Day 30.

Index	Name	P-value	Adjusted p-value
1	histone H3-K9 methylation (GO:0051567)	0.001399	0.02099
2	granulocyte differentiation (GO:0030851)	0.003994	0.02518
3	regulation of protein transport (GO:0051223)	0.005589	0.02518
4	histone lysine methylation (GO:0034968)	0.006783	0.02518
5	regulation of double-strand break repair (GO:2000779)	0.008374	0.02518
6	negative regulation of proteolysis (GO:0045861)	0.008573	0.02518
7	regulation of macromolecule metabolic process (GO:0060255)	0.0005892	0.01767
8	protein deacetylation (GO:0006476)	0.008970	0.02518
9	regulation of DNA repair (GO:0006282)	0.008970	0.02518
10	histone deacetylation (GO:0016575)	0.009368	0.02518

Table S12. GO:BP terms for genes with both TF and Epifactor roles with their altered expression at the terminal differentiation (Day 46).

Index	Name	P-value	Adjusted p-value
1	negative regulation of transcription by RNA polymerase II (GO:0000122)	0.00003983	0.0005576
2	negative regulation of transcription, DNA-templated (GO:0045892)	0.0001062	0.0007432
3	regulation of transcription by RNA polymerase II (GO:0006357)	0.001340	0.006255
4	negative regulation of epidermal growth factor-activated receptor activity (GO:0007175)	0.001949	0.006821
5	negative regulation of protein tyrosine kinase activity (GO:0061099)	0.003596	0.008089
6	negative regulation of signaling receptor activity (GO:2000272)	0.003895	0.008089
7	regulation of epidermal growth factor-activated receptor activity (GO:0007176)	0.004045	0.008089
8	histone exchange (GO:0043486)	0.005689	0.009220
9	ATP-dependent chromatin remodeling (GO:0043044)	0.006287	0.009220
10	negative regulation of epidermal growth factor receptor signaling pathway (GO:0042059)	0.006586	0.009220

Table S13. GO:BP terms for *EZH2* and *VDR* genes.

Index	Name	P-value	Adjusted p-value
1	negative regulation of gene expression (GO:0010629)	0.0002584	0.005954
2	positive regulation of multicellular organismal process (GO:0051240)	0.0002967	0.005954
3	negative regulation of nucleic acid-templated transcription (GO:1903507)	0.0005371	0.005954
4	negative regulation of cellular macromolecule biosynthetic process (GO:2000113)	0.0007467	0.005954
5	cellular response to vitamin D (GO:0071305)	0.0004999	0.005954
6	negative regulation of retinoic acid receptor signaling pathway (GO:0048387)	0.0006999	0.005954
7	regulation of calcidiol 1-monooxygenase activity (GO:0060558)	0.0006999	0.005954
8	negative regulation of keratinocyte proliferation (GO:0010839)	0.0007998	0.005954
9	chromatin silencing at telomere (GO:0006348)	0.0007998	0.005954
10	negative regulation of transcription by RNA polymerase II (GO:0000122)	0.001168	0.007176

Table S14. GO:BP terms for genes with both TF and Epifactor roles with only similar altered expression in both late stages (D30 and D46).

Index	Name	P-value	Adjusted p-value
1	positive regulation of nucleic acid-templated transcription (GO:1903508)	2.611e-8	0.000005572
2	regulation of transcription, DNA-templated (GO:0006355)	3.678e-8	0.000005572
3	ATP-dependent chromatin remodeling (GO:0043044)	9.794e-8	0.000007419
4	positive regulation of transcription, DNA-templated (GO:0045893)	9.773e-8	0.000007419
5	regulation of transcription by RNA polymerase II (GO:0006357)	3.557e-7	0.00002156
6	chromatin remodeling (GO:0006338)	0.000003711	0.0001874
7	chromatin organization (GO:0006325)	0.00001327	0.0005744
8	negative regulation of transcription by RNA polymerase II (GO:0000122)	0.00005471	0.002072
9	negative regulation of nucleic acid-templated transcription (GO:1903507)	0.00009843	0.003314
10	negative regulation of cellular macromolecule biosynthetic process (GO:2000113)	0.0002126	0.006442

Table S15. GO:BP terms for genes with both TF and Epifactor roles with their overexpression from Day 7 toward terminal differentiation (D30 and D46)

Index	Name	P-value	Adjusted p-value
1	regulation of transcription by RNA polymerase II (GO:0006357)	1.759e-7	0.00001077
2	regulation of transcription, DNA-templated (GO:0006355)	2.013e-7	0.00001077
3	negative regulation of transcription, DNA-templated (GO:0045892)	8.293e-7	0.00002958
4	negative regulation of transcription by RNA polymerase II (GO:0000122)	0.000005182	0.0001386
5	regulation of macromolecule metabolic process (GO:0060255)	0.00007916	0.001694
6	negative regulation of nucleic acid-templated transcription (GO:1903507)	0.0009390	0.01435
7	regulation of gene expression (GO:0010468)	0.0008530	0.01435
8	response to lipid (GO:0033993)	0.001129	0.01511
9	negative regulation of cellular macromolecule biosynthetic process (GO:2000113)	0.001511	0.01797
10	chromatin organization (GO:0006325)	0.001744	0.01866

Table S16. GO:BP terms for the genes with only Epifactor role with significant altered expression in all stages of muscle differentiation.

Index	Name	P-value	Adjusted p-value
1	negative regulation of transcription, DNA-templated (GO:0045892)	2.502e-8	0.000002994
2	regulation of transcription by RNA polymerase II (GO:0006357)	1.836e-8	0.000002994
3	regulation of transcription, DNA-templated (GO:0006355)	2.375e-8	0.000002994
4	chromatin organization (GO:0006325)	1.315e-7	0.00001180
5	chromatin remodeling (GO:0006338)	8.117e-7	0.00005828
6	regulation of nucleic acid-templated transcription (GO:1903506)	0.000006585	0.0003940
7	regulation of cellular macromolecule biosynthetic process (GO:2000112)	0.00001143	0.0005862
8	negative regulation of gene expression, epigenetic (GO:0045814)	0.00001766	0.0006433
9	negative regulation of transcription by RNA polymerase II (GO:0000122)	0.00001482	0.0006433
10	positive regulation of transcription, DNA-templated (GO:0045893)	0.00001792	0.0006433

Table S17. GO:BP terms for *DTX3L* gene.

Index	Name	P-value	Adjusted p-value
1	regulation of receptor catabolic process (GO:2000644)	0.0002500	0.003143
2	positive regulation of receptor catabolic process (GO:2000646)	0.0003500	0.003143
3	positive regulation of protein localization to early endosome (GO:1902966)	0.0004000	0.003143
4	positive regulation of protein localization to endosome (GO:1905668)	0.0004000	0.003143
5	regulation of protein localization to early endosome (GO:1902965)	0.0004000	0.003143
6	positive regulation of chromatin binding (GO:0035563)	0.0004500	0.003143
7	histone H2B ubiquitination (GO:0033523)	0.0005000	0.003143
8	regulation of chromatin binding (GO:0035561)	0.0006000	0.003300
9	negative regulation of ubiquitin-protein transferase activity (GO:0051444)	0.0007500	0.003400
10	positive regulation of double-strand break repair via nonhomologous end joining (GO:2001034)	0.0008000	0.003400

Table S18. GO:BP terms for the genes with only Epifactor role with significant altered expression at Day 2.

Index	Name	P-value	Adjusted p-value
1	positive regulation of blood vessel endothelial cell migration (GO:0043536)	0.000004635	0.001270
2	positive regulation of cell migration involved in sprouting angiogenesis (GO:0090050)	0.00006917	0.006318
3	histone modification (GO:0016570)	0.00006271	0.006318
4	regulation of cell migration involved in sprouting angiogenesis (GO:0090049)	0.0002230	0.01527
5	negative regulation of intracellular signal transduction (GO:1902532)	0.0003210	0.01759
6	protein deacetylation (GO:0006476)	0.0004428	0.01860
7	cellular protein modification process (GO:0006464)	0.0004753	0.01860
8	negative regulation of gene expression (GO:0010629)	0.001319	0.04518
9	negative regulation of tubulin deacetylation (GO:1904428)	0.003495	0.04673
10	regulation of adenylate cyclase-activating G protein-coupled receptor signaling pathway (GO:0106070)	0.003495	0.04673

Table S19.GO:BP terms for the genes with only Epifactor role with significant altered expression at Day 7.

Index	Name	P-value	Adjusted p-value
1	positive regulation of gene expression, epigenetic (GO:0045815)	0.0002836	0.01142
2	DNA repair (GO:0006281)	0.0002574	0.01142
3	cellular response to DNA damage stimulus (GO:0006974)	0.0004127	0.01142
4	chromatin remodeling (GO:0006338)	0.0009235	0.01791
5	histone modification (GO:0016570)	0.001129	0.01791
6	regulation of transcription by RNA polymerase II (GO:0006357)	0.001397	0.01791
7	regulation of transcription, DNA-templated (GO:0006355)	0.001511	0.01791
8	chromatin organization (GO:0006325)	0.001744	0.01809
9	cerebral cortex neuron differentiation (GO:0021895)	0.002248	0.01923
10	double-strand break repair (GO:0006302)	0.002317	0.01923

Table S20.GO:BP terms for the genes with only Epifactor role with significant altered expression at Day 30.

Index	Name	P-value	Adjusted p-value
1	positive regulation of histone H3-K4 methylation (GO:0051571)	2.306e-7	0.00006224
2	regulation of histone H3-K4 methylation (GO:0051569)	4.901e-7	0.00006224
3	positive regulation of peptidyl-lysine acetylation (GO:2000758)	5.763e-7	0.00006224
4	positive regulation of histone methylation (GO:0031062)	0.000001310	0.0001061
5	chromosome organization (GO:0051276)	0.000001682	0.0001090
6	DNA replication-dependent nucleosome assembly (GO:0006335)	0.00003048	0.001234
7	DNA replication-dependent nucleosome organization (GO:0034723)	0.00003048	0.001234
8	response to ionizing radiation (GO:0010212)	0.00002934	0.001234
9	chromatin assembly or disassembly (GO:0006333)	0.00004466	0.001554
10	negative regulation of histone methylation (GO:0031061)	0.00005275	0.001554

Table S21. GO:BP terms for the genes with only Epifactor role with significant altered expression at Day 46.

Index	Name	P-value	Adjusted p-value
1	DNA metabolic process (GO:0006259)	7.280e-8	0.00002082
2	base-excision repair, gap-filling (GO:0006287)	0.000003283	0.0003129
3	histone modification (GO:0016570)	0.000002250	0.0003129
4	regulation of DNA metabolic process (GO:0051052)	0.000007570	0.0004928
5	base-excision repair (GO:0006284)	0.000008616	0.0004928
6	regulation of histone modification (GO:0031056)	0.00005275	0.002155
7	DNA deamination (GO:0045006)	0.00005275	0.002155
8	DNA repair (GO:0006281)	0.00009864	0.003526
9	regulation of mRNA catabolic process (GO:0061013)	0.0001415	0.004495
10	DNA modification (GO:0006304)	0.0002184	0.005729

Table S22. GO:BP terms for the genes with only Epifactor role with significant altered expression at Days 2 and 7.

Index	Name	P-value	Adjusted p-value
1	chromatin remodeling (GO:0006338)	1.360e-7	0.00001319
2	chromatin organization (GO:0006325)	0.00004054	0.001966
3	regulation of transcription by RNA polymerase II (GO:0006357)	0.0002532	0.008188
4	chromatin assembly (GO:0031497)	0.0005802	0.01407
5	negative regulation of transcription, DNA-templated (GO:0045892)	0.0008371	0.01624
6	chromosome organization (GO:0051276)	0.001218	0.01969
7	regulation of DNA topoisomerase (ATP-hydrolyzing) activity (GO:2000371)	0.002498	0.02642
8	positive regulation of DNA topoisomerase (ATP-hydrolyzing) activity (GO:2000373)	0.002498	0.02642
9	maintenance of DNA methylation (GO:0010216)	0.002997	0.02642
10	chromatin organization involved in negative regulation of transcription (GO:0097549)	0.002997	0.02642

Table S23. Epifactor genes with expression alteration from D7 toward late stages.

GeneID	D7	D30	D46
TET1	↓	↓	↓
AURKC	↓	↓	↓
BCOR	↓	↓	↓
CARM1	↓	↓	↓
CDYL	↓	↓	↓
EXOSC5	↓	↓	↓
EXOSC7	↓	↓	↓
MYBBP1A	↓	↓	↓
FBL	↓	↓	↓
HMG5	↓	↓	↓
KAT7	↓	↓	↓
KDM2B	↓	↓	↓
KDM6B	↑	↑	↑
KMT2C	↑	↑	↑
NEK9	↑	↑	↑
PHF2	↑	↑	↑
PRDM11	↑	↑	↑
PRMT2	↑	↑	↑
PRR14	↑	↑	↑
RING1	↑	↑	↑
RPS6KA3	↑	↑	↑
RPS6KA5	↑	↑	↑
SIRT2	↑	↑	↑
SPOP	↑	↑	↑
ARRB1	↑	↑	↑
BCORL1	↑	↑	↑
ERCC6	↑	↑	↑
CBX4	↑	↑	↑
EYA1	↑	↑	↑
EYA4	↑	↑	↑
HCFC2	↑	↑	↑
ASXL3	↑	↑	↑
KAT2B	↑	↑	↑

Table S24. The over- and lower expression of some Epifactor genes found at both D30 and D46.

GeneID	D30	D46	GeneID	D30	D46
ACTL6A	↓	↓	MASTL	↓	↓
ACTR3B	↓	↓	MCRS1	↓	↓
ACTR5	↓	↓	MDC1	↓	↓
ANP32E	↓	↓	MSH6	↓	↓
AURKA	↓	↓	MTF2	↓	↓
AURKB	↓	↓	NASP	↓	↓
BRD7	↓	↓	NAT10	↓	↓
BUB1	↓	↓	NCL	↓	↓
CDC6	↓	↓	NOC2L	↓	↓
CDK7	↓	↓	NPM1	↓	↓
CECR2	↓	↓	PARP1	↓	↓
CHAF1A	↓	↓	PCNA	↓	↓
CHAF1B	↓	↓	PELP1	↓	↓
CHEK1	↓	↓	PPM1G	↓	↓
CIT	↓	↓	PRMT1	↓	↓
DAXX	↓	↓	PRMT5	↓	↓
DDX21	↓	↓	RAD54B	↓	↓
DNMT1	↓	↓	RAD54L	↓	↓
DNTTIP2	↓	↓	RCC1	↓	↓
DPY30	↓	↓	RNF168	↓	↓
EXOSC1	↓	↓	RUVBL1	↓	↓
EXOSC2	↓	↓	RUVBL2	↓	↓
EXOSC3	↓	↓	SENP3	↓	↓
EXOSC8	↓	↓	SET	↓	↓
EXOSC9	↓	↓	SETD1A	↓	↓
HCFC1	↓	↓	SETD6	↓	↓
HDGF	↓	↓	SF3B3	↓	↓
HELLS	↓	↓	SFMBT1	↓	↓
HIRA	↓	↓	SMARCA4	↓	↓
HJURP	↓	↓	SMARCD1	↓	↓
INO80C	↓	↓	SRSF3	↓	↓
JADE3	↓	↓	SS18L2	↓	↓
KDM1A	↓	↓	SUPT16H	↓	↓
KDM6A	↓	↓	SUV39H1	↓	↓
LAS1L	↓	↓	TRIM28	↓	↓

MAPKAPK3	↓	↓	TTK	↓	↓
TAF5	↓	↓	UBE2T	↓	↓
TAF6	↓	↓	USP36	↓	↓
TEX10	↓	↓	USP7	↓	↓
TLK1	↓	↓	VRK1	↓	↓
TONSL	↓	↓	WDR77	↓	↓
TRIM24	↓	↓	YWHAE	↓	↓
ZMYND8	↓	↓	GADD45B	↑	↑
ATN1	↑	↑	MSL1	↑	↑
ZMYND11	↑	↑	MST1	↑	↑
SMYD1	↑	↑	KMT2E	↑	↑
RPS6KA4	↑	↑	HDAC6	↑	↑
PIWIL4	↑	↑	L3MBTL4	↑	↑
NEK6	↑	↑	MBD6	↑	↑
CBX6	↑	↑	MBD	↑	↑
APOBEC2	↑	↑			

Table S25. Genes with significant OS in TCGA-BRCA.

	symbol	HR	lower95	upper95	pValue
ENSG00000136274	NACAD	0,47768679	0,34676158	0,65804484	7,22E-06
ENSG00000267194	AC002546.1	0,50573422	0,36657491	0,69772125	2,69E-05
ENSG00000179958	DCTPP1	1,85964878	1,34812159	2,56526831	0,000121
ENSG00000174514	MFSD4A	0,5548634	0,40248433	0,76493261	0,000282
ENSG00000083307	GRHL2	1,81833042	1,32161278	2,50173542	0,000319
ENSG00000171462	DLK2	0,5646934	0,40931749	0,77904963	4,00E-04
ENSG00000111796	KLRB1	0,56989066	0,41410566	0,7842814	0,00058
ENSG00000147676	MAL2	1,77423398	1,28964046	2,4409177	0,000601
ENSG00000272398	CD24	1,77404005	1,28949936	2,44065116	0,000609
ENSG00000130054	FAM155B	1,7509393	1,27271546	2,40885611	0,000781
ENSG00000159214	CCDC24	0,5774295	0,41967965	0,7944746	0,000886
ENSG00000123989	CHPF	1,70523957	1,23482164	2,35486802	0,000926
ENSG00000091513	TF	0,58585091	0,42473916	0,80807544	0,000947
ENSG00000095637	SORBS1	0,58239845	0,42330404	0,80128685	0,00106
ENSG00000120332	TNN	0,5882474	0,42745584	0,80952223	0,00127
ENSG00000082458	DLG3	1,69715674	1,23358216	2,33494055	0,00142
ENSG00000123095	BHLHE41	0,59336823	0,43100096	0,81690274	0,00143
ENSG00000177119	ANO6	1,67910692	1,21968874	2,31157341	0,00147
ENSG00000008283	CYB561	1,67168948	1,21303876	2,30375633	0,00155
ENSG00000175920	DOK7	0,59222571	0,43047436	0,81475538	0,0016

ENSG00000198624	CCDC69	0,59860846	0,43503184	0,82369163	0,00175
ENSG00000196878	LAMB3	0,60394158	0,43848436	0,83183224	0,00178
ENSG00000095203	EPB41L4B	1,65894814	1,20579412	2,28240369	0,00221
ENSG00000168824	HGNC:18790	0,60579602	0,44011732	0,83384316	0,00224
ENSG00000167553	TUBA1C	1,65145435	1,20005575	2,27264566	0,00225
ENSG00000006625	GGCT	1,65233646	1,20088315	2,27350662	0,00228
ENSG00000102100	SLC35A2	1,63406375	1,18565104	2,25206596	0,00251
ENSG00000172172	MRPL13	1,64200825	1,19340953	2,25923376	0,00258
ENSG00000132465	JCHAIN	0,61317746	0,44492666	0,84505298	0,00259
ENSG00000171346	KRT15	0,6121907	0,44469056	0,84278257	0,00261
ENSG00000198885	ITPRIPL1	0,61343513	0,44571445	0,84426847	0,00262
ENSG00000139714	MORN3	0,61167041	0,44450877	0,84169475	0,00277
ENSG00000111087	GLI1	0,61249221	0,44520017	0,84264727	0,00278
ENSG00000105559	PLEKHA4	0,61182267	0,44462008	0,84190301	0,00282
ENSG00000173272	MZT2A	0,6129865	0,44552901	0,84338491	0,00288
ENSG00000174788	PCP2	0,61186349	0,4447425	0,84178358	0,00288
ENSG00000103269	RHBDL1	0,61033231	0,44361459	0,83970531	0,00296
ENSG00000105173	CCNE1	1,63059486	1,18499253	2,24376065	0,00297
ENSG00000134827	TCN1	0,6148354	0,44682214	0,84602472	0,003
ENSG00000240038	AMY2B	0,61614329	0,44769752	0,84796664	0,00311
ENSG00000128422	KRT17	0,61870451	0,44933966	0,85190625	0,00312
ENSG00000140538	NTRK3	0,62248879	0,45083293	0,85950308	0,00325
ENSG00000244509	APOBEC3C	0,61885788	0,44981025	0,85143697	0,00355
ENSG00000071553	ATP6AP1	1,60386837	1,16409168	2,20978621	0,00364
ENSG00000139737	SLAIN1	0,62308199	0,45255511	0,85786494	0,00373
ENSG00000144331	ZNF385B	0,62678539	0,45495874	0,86350672	0,00405
ENSG00000076554	TPD52	1,60030649	1,1629183	2,20220189	0,00418
ENSG00000181617	FDCSP	0,6282241	0,45601858	0,8654593	0,00423
ENSG00000197275	RAD54B	1,59487874	1,15899903	2,19468537	0,00448
ENSG00000186847	KRT14	0,6302628	0,45752013	0,8682267	0,0045
ENSG00000241294	IGKV2-24	0,6275659	0,45611829	0,86345793	0,00454
ENSG00000156170	NDUFAF6	1,59721628	1,16097767	2,19737202	0,00469
ENSG00000113396	SLC27A6	0,63145331	0,45839684	0,86984302	0,00472
ENSG00000103257	SLC7A5	1,59195282	1,15704419	2,19033447	0,00472
ENSG00000162595	DIRAS3	0,63184936	0,45868846	0,87038076	0,00475
ENSG00000137726	FXVD6	0,62870188	0,45694713	0,86501486	0,00475
ENSG00000178363	CALML3	0,63229314	0,45917882	0,87067302	0,00478
ENSG00000149043	SYT8	0,63089376	0,45840197	0,86829239	0,0048
ENSG00000144229	THSD7B	0,63294167	0,45965547	0,8715553	0,00501
ENSG00000104413	ESRP1	1,58632482	1,15296892	2,18256225	0,00507
ENSG00000064042	LIMCH1	1,58197202	1,1496747	2,17682051	0,00521
ENSG00000152580	IGSF10	0,63417517	0,46080462	0,87277368	0,00525
ENSG00000162896	PIGR	0,63498789	0,46115915	0,87433941	0,00527
ENSG00000243649	CFB	0,6386172	0,46348846	0,8799182	0,00563

ENSG00000145824	CXCL14	0,63782197	0,46308494	0,87849298	0,00571
ENSG00000147255	IGSF1	0,63914467	0,46387724	0,88063364	0,00578
ENSG00000184557	SOCS3	0,64250681	0,46566755	0,88650153	0,00584
ENSG00000116729	WLS	0,6397986	0,46453935	0,88117884	0,00594
ENSG00000164946	FREM1	0,64032671	0,46537823	0,88104313	0,00631
ENSG00000150627	WDR17	0,64085239	0,46559246	0,88208429	0,00637
ENSG00000108984	MAP2K6	0,64202723	0,46617884	0,8842078	0,00644
ENSG00000138755	CXCL9	0,64111863	0,46595636	0,88212786	0,00652
ENSG00000230937	MIR205HG	0,64571533	0,46829431	0,89035522	0,00661
ENSG00000179388	EGR3	0,64504499	0,46779889	0,88944852	0,00676
ENSG00000105825	TFPI2	0,6441002	0,46770348	0,88702583	0,00692
ENSG00000069399	BCL3	0,64759856	0,47027574	0,89178299	0,00697
ENSG00000136982	DSCC1	1,56279668	1,13582864	2,15026578	0,00719
ENSG00000211890	IGHA2	0,64594835	0,46906249	0,88953877	0,00727
ENSG00000272674	PCDHB16	1,54314286	1,12061485	2,12498513	0,00741
ENSG00000185479	KRT6B	0,64555235	0,46903636	0,88849794	0,00742
ENSG00000081041	CXCL2	0,64798444	0,4705594	0,89230782	0,00758
ENSG00000135318	NT5E	1,54104505	1,11911005	2,12206103	0,00759
ENSG00000160307	S100B	0,64818459	0,47070654	0,89258005	0,00762
ENSG00000186891	TNFRSF18	0,64368395	0,46785985	0,88558364	0,00782
ENSG00000157617	C2CD2	1,54422431	1,12236676	2,12464302	0,00803
ENSG00000154839	SKA1	1,54884599	1,12581712	2,1308291	0,00803
ENSG00000187699	C2orf88	0,6508779	0,47232779	0,8969238	0,00807
ENSG00000129195	PIMREG	1,542571	1,12116943	2,12236013	0,00836
ENSG00000157303	SUSD3	0,65000975	0,47238618	0,89442219	0,00848
ENSG00000102384	CENPI	1,53827056	1,11812322	2,11629297	0,00899
ENSG00000211943	IGHV3-15	0,65419714	0,4753662	0,9003036	0,00925
ENSG00000143127	ITGA10	0,65414095	0,47532508	0,90022683	0,00932
ENSG00000206053	JPT2	1,53265858	1,11398855	2,10867727	0,00934
ENSG00000196218	RYR1	0,65460867	0,47556105	0,90106731	0,00944
ENSG00000175063	UBE2C	1,5315361	1,11322267	2,10703833	0,00955
ENSG00000261578	AP003119.3	1,53207741	1,11362595	2,10776446	0,00956
ENSG00000115884	SDC1	1,52092189	1,10387644	2,09552747	0,0098
ENSG00000251442	LINC01094	1,51374794	1,09827421	2,08639409	0,0101
ENSG00000206075	SERPINB5	0,66064879	0,47882848	0,91150972	0,0101
ENSG00000101194	SLC17A9	0,6545209	0,47574772	0,90047223	0,0104
ENSG00000173432	SAA1	0,66048962	0,47941281	0,90996012	0,0106
ENSG00000211895	IGHA1	0,66017705	0,47951671	0,90890208	0,0107
ENSG00000111859	NEDD9	0,65932361	0,47901738	0,90749863	0,0107
ENSG00000224650	IGHV3-74	0,65799091	0,47826114	0,90526284	0,0108
ENSG00000101003	GINS1	1,5148851	1,10100654	2,08434444	0,0114
ENSG00000051180	RAD51	1,51387313	1,10039808	2,08271161	0,0117
ENSG00000103550	KNOP1	1,50745748	1,0952677	2,07476951	0,0119
ENSG00000162373	BEND5	0,66388279	0,48245102	0,91354427	0,0124

ENSG00000253125	AC055854.1	0,66435321	0,48285915	0,91406612	0,0125
ENSG00000260196	AC124798.1	0,66845333	0,48544804	0,92044838	0,0125
ENSG00000186871	ERCC6L	1,50841349	1,09642844	2,07520269	0,0127
ENSG00000169764	UGP2	1,50994834	1,09750008	2,07739755	0,0127
ENSG00000241755	IGKV1-9	0,66788406	0,48517286	0,91940247	0,0133
ENSG00000085999	RAD54L	1,50213144	1,09184944	2,06658424	0,0133
ENSG00000196091	MYBPC1	0,66976965	0,48605143	0,92292988	0,0134
ENSG00000214796	AC098934.1	1,49915834	1,08969132	2,06248842	0,0136
ENSG00000148053	NTRK2	0,67019409	0,48655952	0,92313498	0,0137
ENSG00000164934	DCAF13	1,49819547	1,08899231	2,06116209	0,0138
ENSG00000164120	HPGD	0,67180354	0,48733061	0,9261064	0,0139
ENSG00000184305	CCSER1	1,50109701	1,09104856	2,06525384	0,0141
ENSG00000240864	IGKV1-16	0,6701957	0,48698066	0,92234109	0,0143
ENSG00000078081	LAMP3	0,67115196	0,48756979	0,9238574	0,0143
ENSG00000090889	KIF4A	1,49342637	1,08552959	2,05459376	0,0148
ENSG00000185950	IRS2	0,67459467	0,48914715	0,93034985	0,0149
ENSG00000227199	ST7-AS1	0,67403486	0,48968365	0,92778876	0,0151
ENSG00000050628	PTGER3	0,675	0,49009628	0,92966429	0,0155
ENSG00000129451	KLK10	0,6740405	0,48979457	0,92759419	0,0157
ENSG00000196074	SYCP2	1,48231809	1,07713954	2,03990924	0,0158
ENSG00000196557	CACNA1H	1,47284427	1,0680908	2,03097922	0,016
ENSG00000151623	NR3C2	0,67526544	0,49083405	0,92899711	0,016
ENSG00000088305	DNMT3B	1,48405179	1,07867418	2,04177475	0,0162
ENSG00000029993	HMGB3	1,47463264	1,06936262	2,03349301	0,0162
ENSG00000211652	IGLV7-43	0,67607584	0,49117981	0,93057274	0,0162
ENSG00000079616	KIF22	1,47934976	1,07499729	2,03579648	0,0162
ENSG00000259207	ITGB3	1,47151564	1,06811873	2,02726367	0,0164
ENSG00000100911	PSME2	0,67675015	0,49185146	0,93115667	0,017
ENSG00000161800	RACGAP1	1,47571758	1,07215396	2,03118436	0,017
ENSG00000182600	C2orf82	0,67729127	0,49229129	0,93181307	0,0177
ENSG00000151748	SAV1	0,68402767	0,49371394	0,94770232	0,0179
ENSG00000115353	TACR1	0,68147964	0,49447289	0,93921125	0,0179
ENSG00000151364	KCTD14	0,68016963	0,49434736	0,9358414	0,0184
ENSG00000004838	ZMYND10	0,67881739	0,49341611	0,93388328	0,0185
ENSG00000156127	BATF	0,6788395	0,49343218	0,9339137	0,0186
ENSG00000102760	RGCC	0,6829155	0,4960687	0,9401391	0,019
ENSG00000196843	ARID5A	0,6844447	0,49686506	0,94284059	0,0191
ENSG00000272143	FGF14-AS2	0,6838943	0,49663738	0,94175637	0,0191
ENSG00000165244	ZNF367	1,46884932	1,06767065	2,02077141	0,0194
ENSG00000011426	ANLN	1,46563155	1,06523436	2,01652888	0,0197
ENSG00000164128	NPY1R	0,68203824	0,49574845	0,93833104	0,0198
ENSG00000214456	PLIN5	0,68141468	0,49530349	0,93745748	0,0198
ENSG00000064270	ATP2C2	1,45737378	1,05757184	2,00831588	0,0199
ENSG00000169129	AFAP1L2	0,68648255	0,49868588	0,94500029	0,02

ENSG00000121904	CSMD2	1,46122269	1,06171015	2,01106842	0,02
ENSG00000117877	CD3EAP	1,46048981	1,06118185	2,01005179	0,0202
ENSG00000272636	DOC2B	0,68676674	0,49837284	0,94637692	0,0203
ENSG00000117245	KIF17	0,68340465	0,49674917	0,94019668	0,0206
ENSG00000113494	PRLR	1,45683254	1,05854516	2,00497921	0,0206
ENSG00000154655	L3MBTL4	0,68615919	0,49856682	0,94433568	0,0208
ENSG00000167705	RILP	0,68652229	0,49883268	0,94483156	0,0208
ENSG00000197506	SLC28A3	0,68722587	0,49908604	0,94628853	0,0209
ENSG00000115648	MLPH	0,6814177	0,49503744	0,93796962	0,0211
ENSG00000099204	ABLIM1	1,46768652	1,06624836	2,02026451	0,0213
ENSG00000099769	IGFALS	0,68454973	0,49758089	0,94177317	0,0213
ENSG00000147576	ADHFE1	0,68853207	0,50004587	0,94806586	0,0215
ENSG00000069535	MAOB	0,68967327	0,500516	0,95031772	0,0215
ENSG00000121211	MND1	1,45576175	1,05815015	2,0027803	0,0223
ENSG00000183092	BEGAIN	0,69059614	0,50139344	0,95119519	0,0225
ENSG00000205221	VIT	0,69015251	0,50149015	0,9497903	0,0226
ENSG00000019102	VSIG2	0,68878551	0,5005753	0,94776047	0,0226
ENSG00000155066	PROM2	1,45187966	1,05533123	1,99743407	0,023
ENSG00000104432	IL7	0,69115341	0,50230499	0,95100198	0,0236
ENSG00000152082	MZT2B	0,69128687	0,50245538	0,95108453	0,0236
ENSG00000158850	B4GALT3	1,44650465	1,05126788	1,99033544	0,024
ENSG00000146410	MTFR2	1,44830601	1,05272463	1,99253467	0,0242
ENSG00000237248	LINC00987	0,69473256	0,50443698	0,95681592	0,0244
ENSG00000158481	CD1C	0,69420854	0,50435127	0,95553541	0,0246
ENSG00000176887	SOX11	1,44340625	1,04902702	1,98605142	0,0246
ENSG00000196584	XRCC2	1,45397851	1,05585992	2,00221022	0,0252
ENSG00000103490	PYCARD	0,69367495	0,5042154	0,95432417	0,0253
ENSG00000182481	KPNA2	1,440097	1,04663322	1,98147671	0,0256
ENSG00000073282	TP63	0,69671318	0,50589388	0,95950805	0,0258
ENSG00000262772	LINC01977	1,4332744	1,04074106	1,97385843	0,0259
ENSG00000166839	ANKDD1A	0,69753373	0,50678824	0,96007221	0,026
ENSG00000085563	ABCB1	0,69750787	0,50663816	0,9602854	0,0261
ENSG00000156970	BUB1B	1,44158331	1,04785174	1,98326002	0,0261
ENSG00000178935	ZNF552	0,69224792	0,50306125	0,9525822	0,0262
ENSG00000163491	NEK10	0,69601584	0,50578024	0,95780342	0,0264
ENSG00000136854	STXBP1	1,43664968	1,04423344	1,97653343	0,0266
ENSG00000117586	TNFSF4	1,43472826	1,04259677	1,97434448	0,0267
ENSG00000198483	ANKRD35	0,6947346	0,50497369	0,95580456	0,0269
ENSG00000198947	DMD	0,69773741	0,50693749	0,96035016	0,027
ENSG00000185022	MAFF	0,69966961	0,5074263	0,96474613	0,027
ENSG00000157152	SYN2	0,697418	0,506806	0,95972	0,0271
ENSG00000185634	SHC4	0,69623475	0,50606074	0,95787478	0,0272
ENSG00000145819	ARHGAP26	0,70041956	0,50843492	0,96489744	0,028
ENSG00000102172	SMS	1,43494164	1,04251207	1,97509224	0,0284

ENSG00000175164	ABO	0,70161264	0,50931362	0,96651703	0,0287
ENSG00000127084	FGD3	0,70053715	0,50898868	0,96417133	0,0288
ENSG00000101447	FAM83D	1,42985517	1,03922169	1,96732404	0,0289
ENSG00000198774	RASSF9	0,70188339	0,50969568	0,96653808	0,0293
ENSG00000235501	AC105942.1	0,69924098	0,50826129	0,96198149	0,0295
ENSG00000128944	KNSTRN	1,42811363	1,0380606	1,96472975	0,0296
ENSG00000116711	PLA2G4A	0,70098655	0,5095296	0,96438392	0,0299
ENSG00000064692	SNCAIP	0,70316472	0,5106376	0,96828088	0,0302
ENSG00000150667	FSIP1	0,70061197	0,50925745	0,96386833	0,0304
ENSG00000244575	IGKV1-27	0,70268353	0,51056081	0,96710153	0,0304
ENSG00000269516	CYP4F23P	0,70359739	0,51110733	0,96858187	0,0305
ENSG00000165092	ALDH1A1	0,70223024	0,51040882	0,96614181	0,0306
ENSG00000215252	GOLGA8B	0,70309898	0,51095934	0,96749025	0,0307
ENSG00000164764	SBSPON	0,70295418	0,5108535	0,96729214	0,031
ENSG00000157456	CCNB2	1,42415857	1,03518426	1,95929139	0,0311
ENSG00000146386	ABRACL	1,42395935	1,03503936	1,95901749	0,0315
ENSG00000138795	LEF1	0,70508547	0,51231972	0,97038139	0,0315
ENSG00000166803	PCLAF	1,42450385	1,03538693	1,95985786	0,0315
ENSG00000139890	REM2	0,70281548	0,5108582	0,96690158	0,0319
ENSG00000134762	DSC3	0,70683996	0,51295738	0,97400436	0,0322
ENSG00000121552	CSTA	0,70713328	0,51370113	0,97340156	0,0325
ENSG00000244513	AC109587.1	0,70859896	0,51347382	0,97787359	0,0327
ENSG00000163659	TIPARP	0,70672028	0,5132507	0,97311812	0,0327
ENSG00000109794	FAM149A	0,70666349	0,51356421	0,97236777	0,0332
ENSG00000144554	FANCD2	1,41657808	1,02960702	1,94898967	0,0332
ENSG00000157992	KRTCAP3	0,70982649	0,51553284	0,97734539	0,0332
ENSG00000171914	TLN2	0,70470359	0,5122295	0,96950127	0,0332
ENSG00000134531	EMP1	1,41108857	1,02534517	1,94195184	0,0333
ENSG00000131435	PDLIM4	0,70731395	0,51395122	0,97342512	0,0333
ENSG00000165071	TMEM71	0,70839017	0,51462292	0,97511521	0,0337
ENSG00000197142	ACSL5	0,70874321	0,51456325	0,97620056	0,0338
ENSG00000101213	PTK6	1,40947042	1,02246293	1,94296225	0,0339
ENSG00000164932	CTHRC1	1,41094937	1,0247264	1,94274113	0,0341
ENSG00000077264	PAK3	0,70952446	0,51513821	0,97726193	0,0341
ENSG00000077274	CAPN6	0,70905597	0,5149668	0,97629666	0,0344
ENSG00000060718	COL11A1	1,40648241	1,0192589	1,94081479	0,0344
ENSG00000137975	CLCA2	0,70913609	0,51537056	0,97575226	0,035
ENSG00000106089	STX1A	1,41103074	1,02558766	1,94133356	0,035
ENSG00000070371	CLTCL1	1,40702853	1,02157248	1,93792347	0,0351
ENSG00000160182	TFF1	0,70719542	0,51403882	0,97293306	0,0351
ENSG00000108018	SORCS1	0,70909857	0,51534316	0,9757009	0,0354
ENSG00000213145	CRIP1	0,7085179	0,51499886	0,97475481	0,036
ENSG00000186081	KRT5	0,7131566	0,51738775	0,98300036	0,0366
ENSG00000177189	RPS6KA3	0,71287722	0,51760519	0,98181769	0,0367

ENSG00000115541	HSPE1	1,40548038	1,02146142	1,93387148	0,037
ENSG00000198756	COLGALT2	0,71491007	0,518682	0,98537525	0,0377
ENSG00000182463	TSHZ2	0,71302062	0,51826648	0,98095945	0,0378
ENSG00000112742	TTK	1,4061814	1,02202555	1,93473258	0,0378
ENSG00000203760	CENPW	1,40459051	1,02094403	1,9324022	0,0379
ENSG00000181885	CLDN7	1,40018876	1,01628858	1,92910616	0,0379
ENSG00000189159	JPT1	1,40519498	1,02138415	1,9332324	0,038
ENSG00000167614	TTYH1	0,71148328	0,51715941	0,97882479	0,038
ENSG00000211829	TRDC	0,71436326	0,51923578	0,98281917	0,0382
ENSG00000270547	LINC01235	1,41032482	1,02445817	1,94152984	0,0384
ENSG00000270959	LPP-AS2	0,71372849	0,51872429	0,98204069	0,0386
ENSG00000152056	AP1S3	1,40159485	1,01865081	1,92850004	0,0389
ENSG00000204338	CYP21A1P	0,7174627	0,52078643	0,98841423	0,039
ENSG00000133048	CHI3L1	0,71672235	0,52001947	0,98783019	0,0391
ENSG00000153446	C16orf89	0,71406475	0,5189698	0,98250125	0,0394
ENSG00000134057	CCNB1	1,39893146	1,01657248	1,92510546	0,0394
ENSG00000221955	SLC12A8	1,39918075	1,01675243	1,92545078	0,0394
ENSG00000122863	CHST3	0,71503249	0,51967628	0,98382682	0,0397
ENSG00000211649	IGLV7-46	0,71591386	0,52024388	0,98517768	0,0402
ENSG00000254389	RHPN1-AS1	1,39750426	1,01578148	1,9226755	0,0402
ENSG00000181788	SIAH2	0,71509251	0,51976705	0,98382015	0,0405
ENSG00000132321	IQCA1	0,7178738	0,52144207	0,98830306	0,0409
ENSG00000137310	TCF19	1,39675241	1,01523619	1,92163885	0,0409
ENSG00000189221	MAOA	0,71655401	0,52083166	0,9858265	0,041
ENSG00000174371	EXO1	1,3959512	1,01465507	1,92053419	0,0413
ENSG00000149925	ALDOA	1,3896415	1,00608	1,91943334	0,0414
ENSG00000144857	BOC	0,71828778	0,52174594	0,98886698	0,0415
ENSG00000211967	IGHV3-53	0,71801318	0,52154438	0,98849292	0,0416
ENSG00000053747	LAMA3	0,71667872	0,52093662	0,98597097	0,0416
ENSG00000086570	FAT2	0,71803441	0,52155997	0,98852185	0,0417
ENSG00000203995	ZYG11A	1,39928992	1,01684728	1,92557164	0,0418
ENSG00000128268	MGAT3	0,71766486	0,52152443	0,98757186	0,0419
ENSG00000115163	CENPA	1,39661643	1,01513664	1,92145313	0,0421
ENSG00000138180	CEP55	1,39709989	1,01539811	1,9222885	0,0421
ENSG00000104332	SFRP1	0,71925383	0,52210578	0,99084532	0,0422
ENSG00000162444	RBP7	0,72037145	0,52271186	0,99277452	0,0426
ENSG00000157766	ACAN	1,38820021	1,00729767	1,91313838	0,0427
ENSG00000169515	CCDC8	0,71825453	0,52202798	0,98824123	0,0427
ENSG00000109846	CRYAB	0,72103471	0,52320115	0,99367339	0,0433
ENSG00000172818	OVOL1	1,38886171	1,00885895	1,91199857	0,0434
ENSG00000140092	FBLN5	0,72113903	0,52368103	0,99305011	0,0437
ENSG00000115257	PCSK4	0,7176556	0,52162832	0,98734968	0,0437
ENSG00000153956	CACNA2D1	0,71933263	0,52281468	0,98971862	0,0438
ENSG00000118946	PCDH17	1,38654024	1,00762402	1,90794761	0,0438

ENSG00000088340	FER1L4	0,71864856	0,52236817	0,98868151	0,0439
ENSG00000166173	LARP6	0,7207567	0,52355791	0,99223067	0,0439
ENSG00000177675	CD163L1	1,38967604	1,01002519	1,91203104	0,044
ENSG00000225698	IGHV3-72	0,7200051	0,52323556	0,9907724	0,044
ENSG00000165309	ARMC3	0,72048344	0,52358525	0,99142667	0,0443
ENSG00000174938	SEZ6L2	1,38749366	1,00812272	1,9096273	0,0443
ENSG00000124343	XG	1,38532148	1,00655712	1,90661372	0,0444
ENSG00000154065	ANKRD29	0,72132795	0,52410474	0,99276723	0,0449
ENSG00000065328	MCM10	1,38945401	1,00995981	1,91154382	0,0449
ENSG00000108829	LRRC59	1,38677761	1,00779551	1,90827617	0,045
ENSG00000151914	DST	0,72216667	0,52481568	0,99372927	0,0453
ENSG00000152583	SPARCL1	0,72084366	0,52395542	0,99171715	0,0454
ENSG00000151322	NPAS3	0,72232829	0,52500019	0,99382469	0,046
ENSG00000198729	PPP1R14C	0,72256387	0,525106	0,99427269	0,0462
ENSG00000159147	DONSON	1,38587909	1,00734744	1,90665185	0,0464
ENSG00000121207	LRAT	0,72418065	0,52607009	0,99689684	0,047
ENSG00000171241	SHCBP1	1,38501853	1,006734	1,90544505	0,047
ENSG00000184992	BRI3BP	1,38031826	1,00242378	1,90067168	0,0472
ENSG00000250305	KIAA1456	0,72481162	0,52619831	0,99839143	0,0472
ENSG00000139597	N4BP2L1	0,72531056	0,52635472	0,99946935	0,0472
ENSG00000197565	COL4A6	0,72528382	0,52687932	0,99840058	0,0476
ENSG00000173894	CBX2	1,38237873	1,00480681	1,90182923	0,048
ENSG00000168542	COL3A1	1,3743095	0,99642903	1,89549536	0,0482
ENSG00000092820	EZR	1,37813581	1,00085765	1,8976308	0,0482
ENSG00000225556	C2CD4D	0,7264943	0,52788767	0,99982251	0,0483
ENSG00000139547	RDH16	1,38171747	1,00432679	1,90091828	0,0485
ENSG00000153064	BANK1	0,72651865	0,52745489	1,00070994	0,0486
ENSG00000166405	RIC3	0,72665058	0,527552	1,00088914	0,0486
ENSG00000129151	BBOX1	0,72661331	0,52752456	1,00083851	0,0487
ENSG00000100558	PLEK2	0,72669791	0,52812692	0,99992982	0,0488
ENSG00000088325	TPX2	1,38075635	1,00363346	1,89958603	0,049
ENSG00000196405	EVL	0,72372966	0,52603126	0,99572909	0,0491
ENSG00000100526	CDKN3	1,37567309	0,99988333	1,89269729	0,0493
ENSG00000129596	CDO1	0,72646731	0,52786791	0,99978564	0,0495
ENSG00000029559	IBSP	1,37340539	0,99628901	1,89326827	0,0495
ENSG00000123096	SSPN	0,72549968	0,52731316	0,99817306	0,0495

Table S26. Genes with prognostic value from TCGA-BRCA across different stages of muscle differentiation.

GeneID	TF_flag	epi_flag	contrast	logFC	Pvalue	Pvalue_adj	significance
ABCB1	not_TF	not_epi	D2_D0	2,97966596	0,00010091	0,00079491	1
ABLIM1	not_TF	not_epi	D2_D0	2,81122339	2,2707E-11	1,281E-08	1
ABO	not_TF	not_epi	D7_D0	4,89581768	0,00445496	0,01273845	1
ABO	not_TF	not_epi	D30_D0	6,06325127	0,00072634	0,00183226	1
ABO	not_TF	not_epi	D46_D0	5,9883875	0,00080699	0,00196647	1
ABRACL	not_TF	not_epi	D30_D0	-2,5904469	2,9558E-06	2,0129E-05	-1
ABRACL	not_TF	not_epi	D46_D0	-2,4572305	4,8457E-06	2,7802E-05	-1
AC098934.1	not_TF	not_epi	D30_D0	-2,7710346	6,8524E-08	1,1321E-06	-1
AC098934.1	not_TF	not_epi	D46_D0	-3,5782845	5,1203E-09	1,5682E-07	-1
ACAN	not_TF	not_epi	D30_D0	5,28235427	0,00183021	0,00405714	1
ACSL5	not_TF	not_epi	D2_D0	2,84828729	0,01059403	0,03218699	1
ACSL5	not_TF	not_epi	D7_D0	2,76281274	0,01276664	0,0301332	1
ACSL5	not_TF	not_epi	D30_D0	3,13874336	0,00485628	0,00933671	1
ACSL5	not_TF	not_epi	D46_D0	5,7051967	2,6103E-06	1,6932E-05	1
AFAP1L2	not_TF	not_epi	D2_D0	1,9363363	0,00130615	0,00597802	1
AFAP1L2	not_TF	not_epi	D7_D0	2,35868522	0,00025372	0,00124321	1
AFAP1L2	not_TF	not_epi	D30_D0	2,7600532	6,1875E-05	0,0002294	1
AFAP1L2	not_TF	not_epi	D46_D0	3,06704578	2,2435E-05	9,6924E-05	1
ALDH1A1	not_TF	not_epi	D2_D0	-6,7209137	6,6459E-06	9,6161E-05	-1
ALDH1A1	not_TF	not_epi	D7_D0	-3,2348745	7,7535E-05	0,00048705	-1
ALDH1A1	not_TF	not_epi	D30_D0	6,30920755	2,8849E-10	2,5609E-08	1
ALDH1A1	not_TF	not_epi	D46_D0	6,24957226	3,1515E-10	2,1784E-08	1
ALDOA	not_TF	not_epi	D7_D0	-2,0030467	0,00424445	0,0122305	-1
AMY2B	not_TF	not_epi	D30_D0	2,15078876	0,00067694	0,00172597	1
AMY2B	not_TF	not_epi	D46_D0	2,13761207	0,00070388	0,00174603	1
ANKDD1A	not_TF	not_epi	D2_D0	-1,4437542	0,00277302	0,01100693	-1
ANKDD1A	not_TF	not_epi	D7_D0	-1,470349	0,0023859	0,00762929	-1
ANKRD29	not_TF	not_epi	D2_D0	-3,3965544	0,00387441	0,01446535	-1
ANKRD29	not_TF	not_epi	D7_D0	-3,2412453	0,00448847	0,01282168	-1
ANKRD35	not_TF	not_epi	D7_D0	-1,4303497	1,4469E-05	0,00013531	-1
ANLN	not_TF	not_epi	D30_D0	-1,0217135	0,00369599	0,00739113	-1
AP003119.3	not_TF	not_epi	D46_D0	-1,0079513	0,00225428	0,00476929	-1
AP1S3	not_TF	not_epi	D2_D0	-2,6101306	0,0008957	0,0044037	-1
AP1S3	not_TF	not_epi	D7_D0	-1,8601282	0,00645042	0,01719851	-1
APOBEC3C	not_TF	epiFactor	D2_D0	-1,0817092	0,00018191	0,00126385	-1
APOBEC3C	not_TF	epiFactor	D7_D0	-1,1356023	0,00011585	0,00066823	-1
ARHGAP26	not_TF	not_epi	D7_D0	-1,247754	1,9428E-06	2,8575E-05	-1
ARHGAP26	not_TF	not_epi	D30_D0	-1,1083883	7,0756E-06	3,9916E-05	-1
ARID5A	not_TF	not_epi	D7_D0	-1,3966405	5,8411E-05	0,00039033	-1
ARID5A	not_TF	not_epi	D46_D0	1,60508737	7,4703E-06	3,9481E-05	1
ATP2C2	not_TF	not_epi	D2_D0	-7,6505531	6,033E-06	8,8585E-05	-1
ATP2C2	not_TF	not_epi	D7_D0	-5,8289776	1,0677E-05	0,00010704	-1
ATP2C2	not_TF	not_epi	D30_D0	-1,5967556	0,00917215	0,01636427	-1
ATP6AP1	not_TF	not_epi	D30_D0	1,24955001	0,00055348	0,00144662	1
BANK1	not_TF	not_epi	D30_D0	1,17556909	0,00724978	0,01329672	1

BANK1	not_TF	not_epi	D46_D0	1,50771127	0,00121203	0,00278063	1
BBOX1	not_TF	not_epi	D30_D0	7,80669219	4,8346E-06	2,976E-05	1
BBOX1	not_TF	not_epi	D46_D0	8,0064974	3,3376E-06	2,0589E-05	1
BCL3	not_TF	not_epi	D7_D0	-2,8727119	0,00051212	0,00216766	-1
BCL3	not_TF	not_epi	D30_D0	2,21867211	0,00021717	0,00065435	1
BCL3	not_TF	not_epi	D46_D0	3,4537656	2,4427E-06	1,6084E-05	1
BEGAIN	not_TF	not_epi	D30_D0	1,63660657	0,00044922	0,00121011	1
BEGAIN	not_TF	not_epi	D46_D0	2,84387655	1,5404E-06	1,1183E-05	1
BEND5	not_TF	not_epi	D2_D0	-1,6563756	0,00144907	0,00651165	-1
BHLHE41	not_TF	not_epi	D2_D0	-4,1138106	8,9609E-06	0,00012101	-1
BHLHE41	not_TF	not_epi	D7_D0	-5,8166981	1,3542E-06	2,2091E-05	-1
BOC	not_TF	not_epi	D2_D0	-3,2594177	0,00265652	0,01062608	-1
BOC	not_TF	not_epi	D7_D0	4,82784717	1,0581E-05	0,00010622	1
BOC	not_TF	not_epi	D30_D0	6,04128031	1,4997E-06	1,203E-05	1
BOC	not_TF	not_epi	D46_D0	7,21034287	3,2919E-07	3,4349E-06	1
BRI3BP	not_TF	not_epi	D30_D0	-1,3387907	2,2757E-05	0,0001016	-1
BUB1B	not_TF	not_epi	D30_D0	-2,8882373	7,1989E-11	9,7541E-09	-1
BUB1B	not_TF	not_epi	D46_D0	-2,6488916	1,9505E-10	1,5647E-08	-1
C16orf89	not_TF	not_epi	D2_D0	-1,3161688	0,01031004	0,03149931	-1
C16orf89	not_TF	not_epi	D7_D0	-2,0019633	0,00075433	0,00297244	-1
C2CD2	not_TF	not_epi	D2_D0	-1,4300019	8,0282E-05	0,00066444	-1
C2CD2	not_TF	not_epi	D30_D0	-1,1321507	0,00061063	0,00157735	-1
C2orf88	not_TF	not_epi	D2_D0	1,68049527	0,00616296	0,0208791	1
C2orf88	not_TF	not_epi	D7_D0	2,17237878	0,00065865	0,00265769	1
C2orf88	not_TF	not_epi	D30_D0	5,27571422	1,3638E-08	3,513E-07	1
C2orf88	not_TF	not_epi	D46_D0	4,48058402	1,1247E-07	1,5067E-06	1
CACNA1H	not_TF	not_epi	D2_D0	-1,5044769	0,0049406	0,01749368	-1
CACNA1H	not_TF	not_epi	D7_D0	-1,6598113	0,0025881	0,00813723	-1
CACNA2D1	not_TF	not_epi	D2_D0	1,108196	0,01136426	0,03401885	1
CACNA2D1	not_TF	not_epi	D30_D0	1,51594459	0,00145013	0,00331544	1
CACNA2D1	not_TF	not_epi	D46_D0	1,60986591	0,00091442	0,00218671	1
CAPN6	not_TF	not_epi	D2_D0	-3,5100209	2,2034E-08	1,4032E-06	-1
CAPN6	not_TF	not_epi	D30_D0	3,60788977	4,3141E-09	1,5185E-07	1
CAPN6	not_TF	not_epi	D46_D0	4,36977216	5,5357E-10	3,3154E-08	1
CBX2	TF	epiFactor	D30_D0	-1,6086227	0,00021053	0,00063742	-1
CBX2	TF	epiFactor	D46_D0	-1,9897066	3,002E-05	0,0001233	-1
CCDC24	not_TF	not_epi	D2_D0	-2,6871471	3,1096E-06	5,4E-05	-1
CCDC24	not_TF	not_epi	D7_D0	-1,4268612	0,00071523	0,002844	-1
CCDC24	not_TF	not_epi	D30_D0	-1,0070682	0,00794748	0,0144146	-1
CCDC69	not_TF	not_epi	D7_D0	-2,0207587	2,1055E-05	0,00018093	-1
CCDC69	not_TF	not_epi	D30_D0	-1,0874055	0,0034832	0,00702342	-1
CCDC69	not_TF	not_epi	D46_D0	-1,5127401	0,00027652	0,00078084	-1
CCDC8	not_TF	not_epi	D30_D0	1,85358195	1,2546E-07	1,7838E-06	1
CCDC8	not_TF	not_epi	D46_D0	2,34291245	8,012E-09	2,1391E-07	1
CCNB1	not_TF	not_epi	D30_D0	-1,931419	1,7574E-07	2,2833E-06	-1
CCNB1	not_TF	not_epi	D46_D0	-2,0103496	1,1075E-07	1,4875E-06	-1
CCNB2	not_TF	not_epi	D30_D0	-2,0388168	1,3008E-09	6,632E-08	-1

CCNB2	not_TF	not_epi	D46_D0	-1,9785687	1,8396E-09	7,7328E-08	-1
CCNE1	not_TF	not_epi	D30_D0	-1,750255	7,582E-08	1,2078E-06	-1
CCNE1	not_TF	not_epi	D46_D0	-1,7285623	8,4391E-08	1,2038E-06	-1
CCSER1	not_TF	not_epi	D2_D0	-2,8380818	0,00769609	0,02494411	-1
CD24	not_TF	not_epi	D2_D0	-1,5614857	0,00153686	0,00682666	-1
CD24	not_TF	not_epi	D7_D0	-1,7878393	0,00054237	0,00227178	-1
CD24	not_TF	not_epi	D30_D0	-3,2070384	2,7299E-06	1,8945E-05	-1
CD24	not_TF	not_epi	D46_D0	-1,6694897	0,00092957	0,00221778	-1
CD3EAP	not_TF	not_epi	D7_D0	-1,0563048	0,00033377	0,00154664	-1
CD3EAP	not_TF	not_epi	D30_D0	-2,6793969	2,3972E-08	5,1913E-07	-1
CD3EAP	not_TF	not_epi	D46_D0	-3,0686224	5,357E-09	1,6176E-07	-1
CDKN3	not_TF	not_epi	D30_D0	-1,6255322	1,0345E-05	5,4098E-05	-1
CDO1	not_TF	not_epi	D30_D0	1,1778521	0,03212063	0,04952179	1
CDO1	not_TF	not_epi	D46_D0	1,70132803	0,00425361	0,00830678	1
CENPA	not_TF	not_epi	D30_D0	-1,6838551	6,8215E-07	6,4624E-06	-1
CENPA	not_TF	not_epi	D46_D0	-1,1052891	5,1312E-05	0,00019012	-1
CENPI	not_TF	not_epi	D7_D0	1,02556585	0,00354184	0,01049291	1
CENPW	not_TF	not_epi	D30_D0	-1,5083476	2,5562E-05	0,00011153	-1
CENPW	not_TF	not_epi	D46_D0	-1,4697292	3,1779E-05	0,0001291	-1
CEP55	not_TF	not_epi	D7_D0	1,41554668	0,00321291	0,00969248	1
CHI3L1	not_TF	not_epi	D2_D0	5,8230364	3,6586E-06	6,1402E-05	1
CHI3L1	not_TF	not_epi	D7_D0	3,73567875	0,00171136	0,00580118	1
CHI3L1	not_TF	not_epi	D30_D0	3,47364274	0,00386134	0,00767528	1
CHI3L1	not_TF	not_epi	D46_D0	11,1105389	2,527E-10	1,9086E-08	1
CHST3	not_TF	not_epi	D7_D0	1,29354586	3,7355E-05	0,00027646	1
CHST3	not_TF	not_epi	D30_D0	1,31073172	3,3033E-05	0,00013744	1
CLDN7	not_TF	not_epi	D2_D0	-5,8816925	2,1584E-08	1,4032E-06	-1
CLDN7	not_TF	not_epi	D7_D0	-8,9835566	2,3355E-09	2,1248E-07	-1
CLDN7	not_TF	not_epi	D30_D0	-4,3246975	2,5198E-07	3,0058E-06	-1
CLDN7	not_TF	not_epi	D46_D0	-2,0769147	0,00018063	0,00054434	-1
COL11A1	not_TF	not_epi	D7_D0	3,69949466	2,2307E-06	3,1793E-05	1
COL11A1	not_TF	not_epi	D30_D0	4,71042279	2,2737E-07	2,7923E-06	1
COL11A1	not_TF	not_epi	D46_D0	3,95297133	1,1902E-06	9,1771E-06	1
COL3A1	not_TF	not_epi	D2_D0	6,93680293	8,0369E-07	1,9148E-05	1
COL3A1	not_TF	not_epi	D7_D0	13,7003944	3,2701E-10	5,4034E-08	1
COL3A1	not_TF	not_epi	D30_D0	17,2828465	3,709E-11	5,8517E-09	1
COL3A1	not_TF	not_epi	D46_D0	18,0335029	2,5342E-11	4,1856E-09	1
COL4A6	not_TF	not_epi	D2_D0	-1,6744312	6,2439E-07	1,5871E-05	-1
COL4A6	not_TF	not_epi	D46_D0	-1,5134133	1,8743E-06	1,3106E-05	-1
COLGALT2	not_TF	not_epi	D2_D0	1,42484828	6,9398E-06	9,9316E-05	1
COLGALT2	not_TF	not_epi	D7_D0	2,80432296	2,6024E-09	2,2755E-07	1
COLGALT2	not_TF	not_epi	D30_D0	1,23884263	3,0532E-05	0,00012881	1
CRYAB	not_TF	not_epi	D2_D0	-1,8171746	0,00121788	0,00564644	-1
CRYAB	not_TF	not_epi	D7_D0	2,46850854	1,3976E-05	0,00013197	1
CRYAB	not_TF	not_epi	D30_D0	2,45333222	1,5146E-05	7,2748E-05	1
CRYAB	not_TF	not_epi	D46_D0	2,80788916	3,6298E-06	2,2057E-05	1
CSMD2	not_TF	not_epi	D2_D0	-2,1013229	0,00013268	0,00098541	-1

CSMD2	not_TF	not_epi	D7_D0	-2,4086331	3,9571E-05	0,00028906	-1
CSMD2	not_TF	not_epi	D46_D0	-1,6579816	0,0008824	0,00212173	-1
CTHRC1	not_TF	not_epi	D2_D0	2,59794596	0,00539131	0,01877777	1
CTHRC1	not_TF	not_epi	D7_D0	5,54659854	3,2434E-06	4,2203E-05	1
CTHRC1	not_TF	not_epi	D30_D0	7,2090538	2,0165E-07	2,5582E-06	1
CTHRC1	not_TF	not_epi	D46_D0	6,53717608	5,5779E-07	5,0964E-06	1
CXCL14	not_TF	not_epi	D7_D0	2,03862481	0,00021977	0,00111401	1
CXCL14	not_TF	not_epi	D30_D0	3,29846052	1,1922E-06	1,0014E-05	1
CXCL14	not_TF	not_epi	D46_D0	6,14367626	9,6166E-10	4,8466E-08	1
CYB561	not_TF	not_epi	D7_D0	-1,7063274	1,6463E-06	2,5478E-05	-1
CYB561	not_TF	not_epi	D30_D0	-1,4343121	1,0403E-05	5,4279E-05	-1
CYP21A1P	not_TF	not_epi	D30_D0	2,49542413	0,00580594	0,01090711	1
CYP21A1P	not_TF	not_epi	D46_D0	3,69183643	0,00013597	0,00042722	1
DCAF13	not_TF	not_epi	D30_D0	-1,4213339	0,00015489	0,00049277	-1
DCAF13	not_TF	not_epi	D46_D0	-1,8629529	1,1619E-05	5,6395E-05	-1
DCTPP1	not_TF	not_epi	D30_D0	-1,5901975	5,9729E-06	3,5016E-05	-1
DCTPP1	not_TF	not_epi	D46_D0	-1,3383624	3,3078E-05	0,00013313	-1
DIRAS3	not_TF	not_epi	D2_D0	3,49734915	0,00018432	0,00127561	1
DIRAS3	not_TF	not_epi	D7_D0	4,97594057	1,6238E-06	2,522E-05	1
DIRAS3	not_TF	not_epi	D30_D0	5,30250588	6,6526E-07	6,3505E-06	1
DIRAS3	not_TF	not_epi	D46_D0	5,34411587	5,8322E-07	5,2683E-06	1
DLG3	not_TF	not_epi	D30_D0	-1,1070631	5,1139E-07	5,2515E-06	-1
DLG3	not_TF	not_epi	D46_D0	-1,2576943	1,1441E-07	1,525E-06	-1
DLK2	not_TF	not_epi	D7_D0	1,98946743	0,01082324	0,02625563	1
DLK2	not_TF	not_epi	D30_D0	2,7046127	0,00089499	0,00219228	1
DLK2	not_TF	not_epi	D46_D0	3,30851314	9,8902E-05	0,00032739	1
DMD	not_TF	not_epi	D7_D0	3,40160642	1,6487E-11	8,009E-09	1
DMD	not_TF	not_epi	D30_D0	3,45071117	1,3984E-11	3,4169E-09	1
DMD	not_TF	not_epi	D46_D0	3,21528591	3,1959E-11	4,9541E-09	1
DNMT3B	not_TF	epiFactor	D2_D0	-2,5540619	7,2044E-10	1,2374E-07	-1
DNMT3B	not_TF	epiFactor	D7_D0	-3,7026189	1,0276E-11	6,4183E-09	-1
DNMT3B	not_TF	epiFactor	D30_D0	-5,8754064	8,5927E-14	3,0229E-10	-1
DNMT3B	not_TF	epiFactor	D46_D0	-6,0502197	6,4638E-14	2,826E-10	-1
DOC2B	not_TF	not_epi	D2_D0	4,07784212	2,7892E-06	4,9722E-05	1
DOK7	not_TF	not_epi	D30_D0	4,48995715	0,00022727	0,00067884	1
DOK7	not_TF	not_epi	D46_D0	4,44053838	0,00025284	0,00072328	1
DSC3	not_TF	not_epi	D2_D0	-1,3284952	0,01608153	0,04482529	-1
DSC3	not_TF	not_epi	D30_D0	3,13775693	1,0692E-05	5,5371E-05	1
DSC3	not_TF	not_epi	D46_D0	1,77489068	0,00154915	0,0034367	1
DSCC1	not_TF	not_epi	D30_D0	-2,5033259	8,4097E-06	4,5859E-05	-1
DSCC1	not_TF	not_epi	D46_D0	-2,2030758	2,6968E-05	0,0001128	-1
EGR3	TF	not_epi	D2_D0	4,70150872	1,0555E-09	1,5642E-07	1
EGR3	TF	not_epi	D7_D0	-2,4851483	5,6322E-05	0,00037882	-1
EGR3	TF	not_epi	D46_D0	3,71869158	1,5446E-08	3,4064E-07	1
EMP1	not_TF	not_epi	D2_D0	-2,8766273	0,00021156	0,00141644	-1
EMP1	not_TF	not_epi	D7_D0	3,62355912	1,1292E-08	6,5387E-07	1
EMP1	not_TF	not_epi	D30_D0	7,10425299	2,2788E-12	1,1721E-09	1

EMP1	not_TF	not_epi	D46_D0	8,31857739	4,0865E-13	4,2038E-10	1
EPB41L4B	not_TF	not_epi	D2_D0	-2,6641864	5,5287E-05	0,00049993	-1
EPB41L4B	not_TF	not_epi	D7_D0	-4,0072105	1,5694E-06	2,4749E-05	-1
EPB41L4B	not_TF	not_epi	D30_D0	-3,0668694	1,6086E-05	7,6421E-05	-1
EPB41L4B	not_TF	not_epi	D46_D0	-2,1227743	0,00036089	0,00098031	-1
ERCC6L	not_TF	not_epi	D30_D0	-2,6732896	2,4448E-07	2,9425E-06	-1
ERCC6L	not_TF	not_epi	D46_D0	-2,0512004	3,5917E-06	2,1878E-05	-1
ESRP1	not_TF	not_epi	D2_D0	-4,5548183	4,1802E-08	2,201E-06	-1
ESRP1	not_TF	not_epi	D7_D0	-9,3737579	2,5982E-10	4,8099E-08	-1
ESRP1	not_TF	not_epi	D30_D0	-4,7303349	2,9842E-08	6,0401E-07	-1
ESRP1	not_TF	not_epi	D46_D0	-3,4811922	5,3588E-07	4,9401E-06	-1
EVL	not_TF	not_epi	D30_D0	1,14746185	7,1867E-05	0,00025956	1
EXO1	not_TF	not_epi	D30_D0	-2,1196427	2,3574E-08	5,1405E-07	-1
EXO1	not_TF	not_epi	D46_D0	-1,6625751	3,5655E-07	3,6507E-06	-1
EZR	not_TF	not_epi	D2_D0	-1,4982402	8,0863E-05	0,00066862	-1
EZR	not_TF	not_epi	D30_D0	-1,8445476	1,0631E-05	5,5134E-05	-1
EZR	not_TF	not_epi	D46_D0	-1,1498342	0,00080859	0,0019689	-1
FAM155B	not_TF	not_epi	D2_D0	-1,5080816	0,00589167	0,02015523	-1
FAM155B	not_TF	not_epi	D7_D0	-4,1980702	2,2014E-06	3,1478E-05	-1
FAM155B	not_TF	not_epi	D30_D0	-3,6822107	6,3137E-06	3,6513E-05	-1
FAM155B	not_TF	not_epi	D46_D0	-5,0959192	5,1923E-07	4,833E-06	-1
FAM83D	not_TF	not_epi	D30_D0	-2,5215431	3,4696E-11	5,6707E-09	-1
FAM83D	not_TF	not_epi	D46_D0	-2,2921784	1,0422E-10	1,0773E-08	-1
FANCD2	not_TF	not_epi	D30_D0	-2,3903971	5,7519E-10	3,9335E-08	-1
FANCD2	not_TF	not_epi	D46_D0	-1,9896945	4,9207E-09	1,5285E-07	-1
FAT2	not_TF	not_epi	D2_D0	-2,4837595	0,00013564	0,00100001	-1
FBLN5	not_TF	not_epi	D7_D0	1,47406417	0,01876548	0,04140433	1
FBLN5	not_TF	not_epi	D30_D0	7,17196941	1,3814E-08	3,5371E-07	1
FBLN5	not_TF	not_epi	D46_D0	6,22514877	5,3663E-08	8,5861E-07	1
FER1L4	not_TF	not_epi	D2_D0	-3,3011313	1,8515E-05	0,00021288	-1
FER1L4	not_TF	not_epi	D7_D0	-2,4282381	0,00017571	0,00093147	-1
FER1L4	not_TF	not_epi	D30_D0	1,87139398	0,00063355	0,00162814	1
FER1L4	not_TF	not_epi	D46_D0	1,70435874	0,00131366	0,00297968	1
FGD3	not_TF	not_epi	D2_D0	-4,7420516	4,2481E-05	0,0004053	-1
FGF14-AS2	not_TF	not_epi	D30_D0	-1,1415726	0,00941448	0,0167478	-1
FGF14-AS2	not_TF	not_epi	D46_D0	-1,9046058	0,00029039	0,00081306	-1
FREM1	not_TF	not_epi	D2_D0	2,21400016	1,402E-09	1,9007E-07	1
FREM1	not_TF	not_epi	D7_D0	3,71804746	2,5607E-12	2,9854E-09	1
FREM1	not_TF	not_epi	D30_D0	3,7596471	2,2547E-12	1,1721E-09	1
FREM1	not_TF	not_epi	D46_D0	2,93496422	4,3043E-11	5,9029E-09	1
FSIP1	not_TF	not_epi	D7_D0	1,71267369	0,00258234	0,00812453	1
FXYD6	not_TF	not_epi	D2_D0	-3,0437663	3,1092E-06	5,4E-05	-1
FXYD6	not_TF	not_epi	D30_D0	2,0763332	4,9764E-05	0,00019203	1
FXYD6	not_TF	not_epi	D46_D0	2,18152237	3,1001E-05	0,00012667	1
GGCT	not_TF	not_epi	D30_D0	-1,4625197	0,00244199	0,00519594	-1
GGCT	not_TF	not_epi	D46_D0	-1,2764485	0,00582948	0,01095604	-1
GIN51	not_TF	not_epi	D30_D0	-2,4998416	5,9859E-10	4,0075E-08	-1

GINS1	not_TF	not_epi	D46_D0	-2,0191768	7,0898E-09	1,9649E-07	-1
GLI1	TF	not_epi	D7_D0	-1,5995005	0,00067622	0,00271668	-1
GLI1	TF	not_epi	D30_D0	-1,1399831	0,00695781	0,01282173	-1
GOLGA8B	not_TF	not_epi	D30_D0	1,76433654	0,0004078	0,00111308	1
GOLGA8B	not_TF	not_epi	D46_D0	1,12971016	0,00988041	0,01743626	1
GRHL2	TF	not_epi	D2_D0	-5,2622976	1,8103E-05	0,00020864	-1
GRHL2	TF	not_epi	D7_D0	-11,767106	1,27E-06	2,1093E-05	-1
GRHL2	TF	not_epi	D30_D0	-4,0652129	9,3818E-05	0,00032412	-1
GRHL2	TF	not_epi	D46_D0	-1,8819812	0,0106509	0,01864868	-1
HMGB3	TF	not_epi	D2_D0	1,31169332	3,0129E-07	9,1713E-06	1
HMGB3	TF	not_epi	D30_D0	-2,1733451	7,4769E-10	4,6699E-08	-1
HMGB3	TF	not_epi	D46_D0	-2,8169915	3,4114E-11	5,143E-09	-1
HPGD	not_TF	not_epi	D2_D0	-4,1459969	7,5455E-06	0,0001065	-1
HPGD	not_TF	not_epi	D7_D0	-2,9765832	5,2837E-05	0,0003601	-1
HPGD	not_TF	not_epi	D30_D0	1,322597	0,00512548	0,00979183	1
HPGD	not_TF	not_epi	D46_D0	2,65513938	1,2782E-05	6,1108E-05	1
HSPE1	not_TF	not_epi	D30_D0	-1,9291207	0,000321	0,00090542	-1
HSPE1	not_TF	not_epi	D46_D0	-1,9282527	0,00031376	0,00086723	-1
IGSF1	not_TF	not_epi	D2_D0	-1,5045367	8,1851E-07	1,9422E-05	-1
IGSF1	not_TF	not_epi	D7_D0	-1,3174227	3,4578E-06	4,4043E-05	-1
IGSF10	not_TF	not_epi	D30_D0	4,1224624	1,9522E-05	8,9537E-05	1
IGSF10	not_TF	not_epi	D46_D0	5,32974713	1,9864E-06	1,3758E-05	1
IL7R	not_TF	not_epi	D30_D0	7,84357873	0,00126629	0,00295521	1
IL7R	not_TF	not_epi	D46_D0	5,80028252	0,01354883	0,02308251	1
IRS2	not_TF	not_epi	D2_D0	-1,1688454	0,00176281	0,00764992	-1
ITGA10	not_TF	not_epi	D7_D0	2,1861882	0,00326504	0,00981419	1
ITGA10	not_TF	not_epi	D30_D0	4,6641999	9,3103E-07	8,2149E-06	1
ITGA10	not_TF	not_epi	D46_D0	5,92115632	5,1033E-08	8,3097E-07	1
ITGB3	not_TF	not_epi	D30_D0	6,62983447	0,00012061	0,00039955	1
ITGB3	not_TF	not_epi	D46_D0	7,40100986	4,2094E-05	0,00016218	1
ITPRIPL1	not_TF	not_epi	D30_D0	-1,0044371	0,00366181	0,00733621	-1
JPT1	not_TF	not_epi	D46_D0	-1,0182998	0,00044193	0,00116551	-1
JPT2	not_TF	not_epi	D30_D0	-1,4597476	3,5811E-05	0,00014704	-1
JPT2	not_TF	not_epi	D46_D0	-1,1881112	0,00024276	0,00069848	-1
KCTD14	not_TF	not_epi	D7_D0	-2,3424762	1,7356E-05	0,00015557	-1
KCTD14	not_TF	not_epi	D30_D0	-1,8428049	0,0001477	0,00047299	-1
KCTD14	not_TF	not_epi	D46_D0	-1,1028419	0,00661451	0,0122381	-1
KIF17	not_TF	not_epi	D2_D0	-3,1352666	4,0612E-07	1,1309E-05	-1
KIF17	not_TF	not_epi	D7_D0	-3,1833498	3,3275E-07	7,5182E-06	-1
KIF17	not_TF	not_epi	D30_D0	-3,5883749	1,2789E-07	1,8069E-06	-1
KIF17	not_TF	not_epi	D46_D0	-4,2340706	2,9566E-08	5,5182E-07	-1
KIF22	not_TF	not_epi	D30_D0	-2,0611001	3,5597E-09	1,3359E-07	-1
KIF22	not_TF	not_epi	D46_D0	-1,8252364	1,4855E-08	3,3264E-07	-1
KIF4A	not_TF	not_epi	D30_D0	-1,3851234	4,3105E-08	7,9528E-07	-1
KIF4A	not_TF	not_epi	D46_D0	-1,1662195	3,2233E-07	3,3831E-06	-1
KLK10	not_TF	not_epi	D2_D0	-5,2720671	9,6679E-07	2,2072E-05	-1
KLK10	not_TF	not_epi	D7_D0	-5,8478226	5,336E-07	1,0738E-05	-1

KLK10	not_TF	not_epi	D30_D0	-2,4698971	0,00015295	0,00048784	-1
KLK10	not_TF	not_epi	D46_D0	-2,0401819	0,0005861	0,00149261	-1
KNOP1	not_TF	not_epi	D30_D0	-2,3264356	1,1505E-10	1,3687E-08	-1
KNOP1	not_TF	not_epi	D46_D0	-2,2694217	1,5374E-10	1,3718E-08	-1
KNSTRN	not_TF	not_epi	D30_D0	-1,0084963	2,0265E-05	9,2412E-05	-1
KNSTRN	not_TF	not_epi	D46_D0	-1,3275424	1,048E-06	8,2818E-06	-1
KPNA2	not_TF	not_epi	D30_D0	-2,0284501	9,8387E-08	1,491E-06	-1
KPNA2	not_TF	not_epi	D46_D0	-2,3550649	1,7773E-08	3,7962E-07	-1
KRT17	not_TF	not_epi	D30_D0	2,26414264	0,01289174	0,02208785	1
KRT17	not_TF	not_epi	D46_D0	4,67521546	4,5707E-05	0,00017343	1
KRTCAP3	not_TF	not_epi	D2_D0	-2,2390615	5,8957E-06	8,6862E-05	-1
KRTCAP3	not_TF	not_epi	D7_D0	-4,1697973	1,6531E-08	8,5693E-07	-1
KRTCAP3	not_TF	not_epi	D30_D0	-3,6074529	6,1665E-08	1,0409E-06	-1
KRTCAP3	not_TF	not_epi	D46_D0	-3,4495095	8,9133E-08	1,2571E-06	-1
L3MBTL4	not_TF	epiFactor	D30_D0	2,3342032	0,00375092	0,00747875	1
L3MBTL4	not_TF	epiFactor	D46_D0	4,36459306	1,7806E-05	7,9965E-05	1
LAMA3	not_TF	not_epi	D2_D0	-3,8652711	0,00020672	0,00138933	-1
LAMA3	not_TF	not_epi	D30_D0	1,53945826	0,01038309	0,0182602	1
LAMA3	not_TF	not_epi	D46_D0	4,01075832	3,6485E-06	2,2139E-05	1
LAMB3	not_TF	not_epi	D2_D0	-3,9684742	0,00088549	0,00436946	-1
LAMB3	not_TF	not_epi	D7_D0	-3,6709997	0,00109483	0,00402684	-1
LAMB3	not_TF	not_epi	D46_D0	4,17850792	7,6653E-06	4,0232E-05	1
LAMP3	not_TF	not_epi	D2_D0	1,62060611	0,00390788	0,01455302	1
LAMP3	not_TF	not_epi	D7_D0	2,29190477	0,00019022	0,00099091	1
LEF1	TF	not_epi	D2_D0	11,6720816	9,5143E-13	1,5126E-09	1
LEF1	TF	not_epi	D7_D0	10,5986541	3,29E-12	3,1965E-09	1
LEF1	TF	not_epi	D30_D0	7,40049789	7,8548E-10	4,8198E-08	1
LEF1	TF	not_epi	D46_D0	7,32459395	9,2332E-10	4,6939E-08	1
LIMCH1	not_TF	not_epi	D2_D0	3,64968174	1,0365E-09	1,5493E-07	1
LIMCH1	not_TF	not_epi	D7_D0	5,64458462	8,432E-12	6,042E-09	1
LIMCH1	not_TF	not_epi	D30_D0	4,54085425	8,7331E-11	1,0987E-08	1
LIMCH1	not_TF	not_epi	D46_D0	5,03064359	2,8508E-11	4,6062E-09	1
LPP-AS2	not_TF	not_epi	D7_D0	2,69348192	5,1839E-06	6,0843E-05	1
LPP-AS2	not_TF	not_epi	D30_D0	2,6755286	5,7051E-06	3,3797E-05	1
LPP-AS2	not_TF	not_epi	D46_D0	3,61483382	1,5273E-07	1,9147E-06	1
LRAT	not_TF	not_epi	D7_D0	-4,6797748	3,14E-06	4,1133E-05	-1
LRAT	not_TF	not_epi	D30_D0	-1,714225	0,00475523	0,00917166	-1
LRAT	not_TF	not_epi	D46_D0	-2,1032424	0,00123775	0,00283396	-1
MAL2	not_TF	not_epi	D2_D0	-2,8258812	0,00013055	0,00097323	-1
MAL2	not_TF	not_epi	D7_D0	-6,5320207	1,4641E-07	4,1363E-06	-1
MAL2	not_TF	not_epi	D30_D0	-3,9912453	7,1595E-06	4,0318E-05	-1
MAL2	not_TF	not_epi	D46_D0	-1,6607083	0,00614938	0,0114869	-1
MAL2	not_TF	not_epi	D46_D0	-1,6607083	0,00614938	0,0114869	-1
MAOA	not_TF	not_epi	D2_D0	2,00461815	4,8893E-06	7,646E-05	1
MAOA	not_TF	not_epi	D7_D0	1,59127284	5,1938E-05	0,0003555	1
MAOB	not_TF	not_epi	D7_D0	-1,4028799	0,0158557	0,03601472	-1
MAOB	not_TF	not_epi	D30_D0	2,12339915	0,00042588	0,00115595	1

MAOB	not_TF	not_epi	D46_D0	1,33556546	0,01148169	0,01995743	1
MAP2K6	not_TF	not_epi	D2_D0	-2,228395	7,931E-05	0,00065827	-1
MCM10	not_TF	not_epi	D30_D0	-2,4616726	4,4473E-09	1,5529E-07	-1
MCM10	not_TF	not_epi	D46_D0	-2,2799684	1,0313E-08	2,5479E-07	-1
MFSD4A	not_TF	not_epi	D2_D0	-4,7341561	0,01457209	0,04142379	-1
MFSD4A	not_TF	not_epi	D30_D0	2,64944571	0,00383176	0,00762082	1
MGAT3	not_TF	not_epi	D30_D0	1,02409957	0,00202026	0,00441077	1
MLPH	not_TF	not_epi	D7_D0	-5,4016022	5,438E-05	0,0003686	-1
MLPH	not_TF	not_epi	D30_D0	-1,7237587	0,02688936	0,04226127	-1
MND1	not_TF	not_epi	D30_D0	-2,1671366	2,1616E-08	4,8548E-07	-1
MND1	not_TF	not_epi	D46_D0	-2,0461959	3,9651E-08	6,8068E-07	-1
MORN3	not_TF	not_epi	D46_D0	1,00210936	0,02883322	0,04520264	1
MRPL13	not_TF	not_epi	D30_D0	-1,1606095	0,00153299	0,00347806	-1
MRPL13	not_TF	not_epi	D46_D0	-1,197447	0,00119266	0,00274292	-1
MTRFR2	not_TF	not_epi	D30_D0	-1,9938414	5,7315E-05	0,00021555	-1
MTRFR2	not_TF	not_epi	D46_D0	-1,740061	0,00017394	0,00052702	-1
MYBPC1	not_TF	not_epi	D30_D0	10,1028254	0,00095497	0,00231726	1
MYBPC1	not_TF	not_epi	D46_D0	12,6207804	9,8358E-05	0,00032597	1
MZT2A	not_TF	not_epi	D7_D0	-1,4907383	0,00177328	0,00597446	-1
N4BP2L1	not_TF	not_epi	D30_D0	2,25866731	0,00271951	0,00568477	1
N4BP2L1	not_TF	not_epi	D46_D0	2,26052275	0,00269489	0,00557401	1
NACAD	not_TF	not_epi	D30_D0	1,71027108	5,0335E-05	0,00019374	1
NACAD	not_TF	not_epi	D46_D0	1,55600574	0,0001232	0,00039266	1
NDUFAF6	not_TF	not_epi	D2_D0	1,19365741	0,00016292	0,00115631	1
NEDD9	not_TF	not_epi	D2_D0	1,9604881	0,0032834	0,01263645	1
NEDD9	not_TF	not_epi	D7_D0	2,03797817	0,00252864	0,00798644	1
NPAS3	TF	not_epi	D2_D0	1,95241983	0,0092413	0,02882219	1
NPAS3	TF	not_epi	D30_D0	5,13950525	7,8909E-07	7,2363E-06	1
NPAS3	TF	not_epi	D46_D0	6,21390125	9,9562E-08	1,3731E-06	1
NPY1R	not_TF	not_epi	D2_D0	-1,5733568	0,00784523	0,0253085	-1
NPY1R	not_TF	not_epi	D7_D0	2,49725118	9,3448E-05	0,00056625	1
NPY1R	not_TF	not_epi	D30_D0	2,56858556	7,2567E-05	0,00026146	1
NR3C2	TF	not_epi	D2_D0	-4,1211832	2,6205E-06	4,7195E-05	-1
NR3C2	TF	not_epi	D7_D0	-1,2591772	0,00524069	0,01449687	-1
NR3C2	TF	not_epi	D30_D0	2,94094612	3,3654E-07	3,7405E-06	1
NR3C2	TF	not_epi	D46_D0	3,15835478	1,4636E-07	1,8521E-06	1
NT5E	not_TF	not_epi	D30_D0	6,31008987	7,9416E-07	7,2676E-06	1
NT5E	not_TF	not_epi	D46_D0	7,33161885	1,8594E-07	2,2196E-06	1
NTRK2	not_TF	not_epi	D7_D0	6,96169255	0,0003131	0,00146954	1
NTRK2	not_TF	not_epi	D30_D0	10,6532041	1,9818E-06	1,4926E-05	1
NTRK2	not_TF	not_epi	D46_D0	10,9984734	1,3905E-06	1,0303E-05	1
NTRK3	not_TF	not_epi	D30_D0	1,90294691	0,02001818	0,03251698	1
OVOL1	TF	not_epi	D2_D0	-4,1453308	0,00118846	0,00553937	-1
OVOL1	TF	not_epi	D7_D0	-6,8096108	0,00025843	0,00126133	-1
PAK3	not_TF	not_epi	D2_D0	-2,3326134	3,4257E-07	1,0001E-05	-1
PAK3	not_TF	not_epi	D30_D0	1,88131778	1,3527E-06	1,1122E-05	1
PAK3	not_TF	not_epi	D46_D0	1,48006896	1,7542E-05	7,8983E-05	1

PCDH17	not_TF	not_epi	D2_D0	1,69107034	0,00152774	0,00679134	1
PCDH17	not_TF	not_epi	D7_D0	3,32767953	1,8324E-06	2,741E-05	1
PCDH17	not_TF	not_epi	D30_D0	4,70004048	3,9599E-08	7,4164E-07	1
PCDH17	not_TF	not_epi	D46_D0	2,94827622	6,9544E-06	3,726E-05	1
PCDHB16	not_TF	not_epi	D30_D0	1,5965822	7,0275E-06	3,9683E-05	1
PCDHB16	not_TF	not_epi	D46_D0	1,75857095	2,4397E-06	1,6076E-05	1
PCLAF	not_TF	not_epi	D2_D0	1,26162054	0,00154294	0,0068467	1
PCSK4	not_TF	not_epi	D2_D0	-1,8385006	0,00048958	0,00271371	-1
PCSK4	not_TF	not_epi	D7_D0	-2,2354543	0,0001031	0,00061205	-1
PCSK4	not_TF	not_epi	D46_D0	-1,014286	0,01986678	0,03242769	-1
PDLIM4	not_TF	not_epi	D2_D0	1,4827976	1,4506E-05	0,00017519	1
PDLIM4	not_TF	not_epi	D7_D0	1,35844423	3,6048E-05	0,00026906	1
PDLIM4	not_TF	not_epi	D30_D0	3,15904472	2,0809E-09	9,2376E-08	1
PDLIM4	not_TF	not_epi	D46_D0	3,18770379	1,851E-09	7,7439E-08	1
PIMREG	not_TF	not_epi	D30_D0	-2,2998768	3,1742E-09	1,2363E-07	-1
PIMREG	not_TF	not_epi	D46_D0	-2,485453	1,2671E-09	5,8933E-08	-1
PLA2G4A	not_TF	not_epi	D7_D0	1,20334294	0,01108669	0,02680435	1
PLA2G4A	not_TF	not_epi	D46_D0	2,02051548	0,00020871	0,00061313	1
PLEKHA4	TF	not_epi	D30_D0	1,00220423	0,01699072	0,028167	1
PLEKHA4	TF	not_epi	D46_D0	2,67896471	7,12E-06	3,8008E-05	1
PPP1R14C	not_TF	not_epi	D2_D0	6,79578642	2,1926E-09	2,6814E-07	1
PPP1R14C	not_TF	not_epi	D7_D0	4,66727817	2,0139E-07	5,2699E-06	1
PPP1R14C	not_TF	not_epi	D30_D0	2,7014221	0,00014429	0,0004641	1
PPP1R14C	not_TF	not_epi	D46_D0	2,96579046	5,0705E-05	0,00018827	1
PRLR	not_TF	not_epi	D30_D0	2,53947668	0,00021721	0,00065435	1
PRLR	not_TF	not_epi	D46_D0	1,75957342	0,00388482	0,00767312	1
PROM2	not_TF	not_epi	D2_D0	-2,9679643	0,00017264	0,00121398	-1
PROM2	not_TF	not_epi	D7_D0	-2,5296063	0,00046534	0,00200587	-1
PROM2	not_TF	not_epi	D46_D0	2,73007008	5,8611E-05	0,00021195	1
PTGER3	not_TF	not_epi	D30_D0	7,20774856	6,2048E-09	2,0132E-07	1
PTGER3	not_TF	not_epi	D46_D0	7,24864424	5,7733E-09	1,6964E-07	1
PYCARD	not_TF	not_epi	D2_D0	-1,2954751	0,01063526	0,03228983	-1
PYCARD	not_TF	not_epi	D7_D0	-3,6595479	6,7957E-06	7,5744E-05	-1
RAD51	TF	epiFactor	D30_D0	-2,0774422	5,7045E-07	5,694E-06	-1
RAD51	TF	epiFactor	D46_D0	-1,2202915	0,00011568	0,00037248	-1
RAD54B	not_TF	epiFactor	D30_D0	-1,1892321	0,01088935	0,01902236	-1
RAD54B	not_TF	epiFactor	D46_D0	-1,3149843	0,00589946	0,01107329	-1
RAD54L	not_TF	epiFactor	D30_D0	-2,4580123	3,2188E-09	1,2509E-07	-1
RAD54L	not_TF	epiFactor	D46_D0	-2,0013542	3,2344E-08	5,9104E-07	-1
RASSF9	not_TF	not_epi	D2_D0	2,8788797	8,4284E-07	1,9785E-05	1
RASSF9	not_TF	not_epi	D7_D0	1,4986452	0,000643	0,00260961	1
RASSF9	not_TF	not_epi	D30_D0	2,447647	5,14E-06	3,1157E-05	1
RASSF9	not_TF	not_epi	D46_D0	3,4967702	9,3413E-08	1,3048E-06	1
RBP7	not_TF	not_epi	D2_D0	-2,3033695	0,00265333	0,01061574	-1
RBP7	not_TF	not_epi	D7_D0	-2,1439337	0,00364379	0,01075304	-1
RDH16	not_TF	not_epi	D2_D0	-2,0694516	0,00296663	0,01163499	-1
RDH16	not_TF	not_epi	D46_D0	-1,4411773	0,01741111	0,02882567	-1

REM2	not_TF	not_epi	D7_D0	-1,844753	0,00822237	0,02094372	-1
RGCC	not_TF	not_epi	D7_D0	7,00449737	0,00011998	0,00068615	1
RGCC	not_TF	not_epi	D30_D0	8,90642335	3,566E-06	2,3409E-05	1
RGCC	not_TF	not_epi	D46_D0	8,99617777	3,0574E-06	1,9144E-05	1
RHBDL1	not_TF	not_epi	D7_D0	-1,6458171	0,00846538	0,02141579	-1
RIC3	not_TF	not_epi	D30_D0	-2,6916337	0,00026257	0,00076595	-1
RIC3	not_TF	not_epi	D46_D0	-3,562487	3,041E-05	0,00012458	-1
RILP	not_TF	not_epi	D2_D0	-1,9712382	0,00035866	0,00213247	-1
RILP	not_TF	not_epi	D7_D0	-2,0646171	0,0002397	0,00119358	-1
RPS6KA3	not_TF	epiFactor	D7_D0	2,89877365	3,0015E-07	6,9987E-06	1
RPS6KA3	not_TF	epiFactor	D30_D0	1,54139924	0,00017527	0,00054578	1
RPS6KA3	not_TF	epiFactor	D46_D0	1,83476479	3,4004E-05	0,00013617	1
RYR1	not_TF	not_epi	D7_D0	-3,6082034	0,00312714	0,00949664	-1
S100B	not_TF	not_epi	D30_D0	6,79693768	0,00133415	0,00308783	1
S100B	not_TF	not_epi	D46_D0	6,09115922	0,00350609	0,00703264	1
SAV1	not_TF	not_epi	D2_D0	-1,0037594	2,112E-05	0,00023615	-1
SAV1	not_TF	not_epi	D30_D0	-1,4018361	5,7361E-07	5,7061E-06	-1
SAV1	not_TF	not_epi	D46_D0	-1,0647172	1,1421E-05	5,565E-05	-1
SBSPON	not_TF	not_epi	D7_D0	5,46164808	1,1099E-07	3,3363E-06	1
SBSPON	not_TF	not_epi	D30_D0	3,23507503	3,589E-05	0,00014722	1
SBSPON	not_TF	not_epi	D46_D0	2,77439066	0,0001704	0,0005179	1
SEZ6L2	not_TF	not_epi	D2_D0	-1,2014522	0,00274076	0,01091072	-1
SEZ6L2	not_TF	not_epi	D7_D0	-1,742552	0,00014426	0,00079436	-1
SEZ6L2	not_TF	not_epi	D46_D0	-1,3640775	0,00108009	0,00251613	-1
SFRP1	not_TF	not_epi	D2_D0	-3,3276506	2,7687E-08	1,6194E-06	-1
SFRP1	not_TF	not_epi	D7_D0	-1,9479983	7,8984E-06	8,4897E-05	-1
SHC4	not_TF	not_epi	D7_D0	1,280908	0,02278613	0,04845967	1
SHC4	not_TF	not_epi	D30_D0	3,12557649	1,1673E-05	5,9185E-05	1
SHC4	not_TF	not_epi	D46_D0	2,64702476	6,9297E-05	0,00024354	1
SKA1	not_TF	not_epi	D2_D0	1,77763284	3,4636E-07	1,0062E-05	1
SKA1	not_TF	not_epi	D30_D0	-1,56105	2,1207E-06	1,5668E-05	-1
SKA1	not_TF	not_epi	D46_D0	-1,0619957	9,9364E-05	0,00032854	-1
SLAIN1	not_TF	not_epi	D2_D0	-2,9013046	6,902E-06	9,912E-05	-1
SLAIN1	not_TF	not_epi	D30_D0	-1,2478892	0,00569769	0,0107309	-1
SLAIN1	not_TF	not_epi	D46_D0	-2,5185389	2,3753E-05	0,00010164	-1
SLC12A8	not_TF	not_epi	D7_D0	-2,105072	0,00950477	0,02354293	-1
SLC17A9	not_TF	not_epi	D7_D0	-4,6127645	0,00015381	0,0008377	-1
SLC17A9	not_TF	not_epi	D46_D0	-2,3002676	0,01012918	0,01783519	-1
SLC27A6	not_TF	not_epi	D7_D0	-1,9561822	0,00036311	0,00164852	-1
SLC7A5	not_TF	not_epi	D30_D0	-3,3640364	5,3793E-06	3,2316E-05	-1
SLC7A5	not_TF	not_epi	D46_D0	-4,3230754	5,1322E-07	4,7918E-06	-1
SMS	not_TF	not_epi	D30_D0	-2,1349086	7,3064E-06	4,0954E-05	-1
SMS	not_TF	not_epi	D46_D0	-2,104009	8,3981E-06	4,3311E-05	-1
SNCAIP	not_TF	not_epi	D2_D0	-2,0326427	0,00013867	0,00101835	-1
SNCAIP	not_TF	not_epi	D30_D0	3,07518741	6,0475E-07	5,9149E-06	1
SNCAIP	not_TF	not_epi	D46_D0	3,00027999	7,8404E-07	6,623E-06	1
SOCS3	not_TF	not_epi	D2_D0	-2,317012	1,229E-05	0,0001544	-1

SOCS3	not_TF	not_epi	D7_D0	-1,2460357	0,00228576	0,00737205	-1
SOCS3	not_TF	not_epi	D46_D0	2,24608403	1,1515E-05	5,5999E-05	1
SORCS1	not_TF	not_epi	D2_D0	-2,6128641	3,7194E-06	6,2125E-05	-1
SORCS1	not_TF	not_epi	D7_D0	-1,0726745	0,00498577	0,01394613	-1
SORCS1	not_TF	not_epi	D30_D0	1,77661679	6,1594E-05	0,0002286	1
SORCS1	not_TF	not_epi	D46_D0	1,08513332	0,0036268	0,00724118	1
SPARCL1	not_TF	not_epi	D2_D0	-3,8993939	0,000713	0,00367275	-1
SPARCL1	not_TF	not_epi	D30_D0	6,79686842	3,3783E-09	1,2985E-07	1
SPARCL1	not_TF	not_epi	D46_D0	7,62015487	1,0904E-09	5,2968E-08	1
SSPN	not_TF	not_epi	D30_D0	3,84457688	5,5253E-07	5,5692E-06	1
SSPN	not_TF	not_epi	D46_D0	4,97996621	3,6239E-08	6,4669E-07	1
ST7-AS1	not_TF	not_epi	D30_D0	1,2176095	0,00302232	0,00621889	1
STXBP1	not_TF	not_epi	D2_D0	-1,2197446	1,0944E-06	2,438E-05	-1
STXBP1	not_TF	not_epi	D7_D0	-1,2111029	1,1785E-06	1,9873E-05	-1
STXBP1	not_TF	not_epi	D30_D0	-1,0062781	8,9882E-06	4,8231E-05	-1
STXBP1	not_TF	not_epi	D46_D0	-1,1897062	1,437E-06	1,0581E-05	-1
SUSD3	not_TF	not_epi	D2_D0	-2,3259078	0,00022235	0,00147177	-1
SUSD3	not_TF	not_epi	D7_D0	-3,2257923	2,1736E-05	0,00018517	-1
SUSD3	not_TF	not_epi	D30_D0	-2,5885537	0,00010936	0,00036826	-1
SUSD3	not_TF	not_epi	D46_D0	-2,1731589	0,00032819	0,00090228	-1
SYCP2	not_TF	not_epi	D7_D0	2,90735087	0,00045712	0,00197433	1
SYCP2	not_TF	not_epi	D30_D0	4,20378584	1,2719E-05	6,3387E-05	1
SYCP2	not_TF	not_epi	D46_D0	3,31727	0,00013406	0,00042197	1
SYN2	not_TF	not_epi	D2_D0	-1,6945072	0,00047636	0,00265642	-1
SYN2	not_TF	not_epi	D7_D0	-3,4111537	1,1711E-06	1,9818E-05	-1
SYN2	not_TF	not_epi	D30_D0	-1,32871	0,00281107	0,00584876	-1
SYN2	not_TF	not_epi	D46_D0	-1,293917	0,00328582	0,00663691	-1
TACR1	not_TF	not_epi	D7_D0	-3,1179657	0,02135103	0,04593732	-1
TACR1	not_TF	not_epi	D30_D0	2,70520621	0,00269463	0,00564154	1
TCF19	TF	not_epi	D2_D0	-1,0587722	2,0132E-05	0,00022759	-1
TF	not_TF	not_epi	D30_D0	7,56329667	1,1733E-06	9,9075E-06	1
TF	not_TF	not_epi	D46_D0	7,22691719	2,1749E-06	1,473E-05	1
TFPI2	not_TF	not_epi	D2_D0	-1,8943314	0,00337426	0,01291508	-1
THSD7B	not_TF	not_epi	D7_D0	-2,5048598	0,01618676	0,03665337	-1
THSD7B	not_TF	not_epi	D30_D0	1,83280946	0,02231985	0,03584951	1
TIPARP	TF	not_epi	D7_D0	2,20920244	2,7606E-06	3,7368E-05	1
TIPARP	TF	not_epi	D30_D0	1,08633867	0,00206781	0,00449942	1
TIPARP	TF	not_epi	D46_D0	1,29888089	0,00049361	0,00128399	1
TNFSF4	not_TF	not_epi	D7_D0	3,68287174	0,00095087	0,00358779	1
TNFSF4	not_TF	not_epi	D30_D0	5,9118135	8,6468E-06	4,6742E-05	1
TNFSF4	not_TF	not_epi	D46_D0	5,42607441	2,0793E-05	9,093E-05	1
TNN	not_TF	not_epi	D2_D0	-4,1617468	0,00022353	0,00147734	-1
TP63	TF	not_epi	D30_D0	4,93245932	8,715E-05	0,00030415	1
TP63	TF	not_epi	D46_D0	5,12792942	5,7164E-05	0,0002077	1
TPD52	not_TF	not_epi	D2_D0	-1,5302801	0,01029356	0,03145444	-1
TPD52	not_TF	not_epi	D7_D0	-3,3271917	3,9037E-05	0,00028624	-1
TPD52	not_TF	not_epi	D30_D0	-2,7426573	0,00018928	0,00058257	-1

TPD52	not_TF	not_epi	D46_D0	-2,1632936	0,00112778	0,00261122	-1
TPX2	not_TF	not_epi	D30_D0	-1,772836	7,9652E-09	2,4178E-07	-1
TPX2	not_TF	not_epi	D46_D0	-1,4582173	8,1378E-08	1,1703E-06	-1
TSHZ2	TF	not_epi	D7_D0	8,31503731	2,3942E-08	1,0989E-06	1
TSHZ2	TF	not_epi	D30_D0	10,5796676	8,2706E-10	4,9622E-08	1
TSHZ2	TF	not_epi	D46_D0	11,4244385	3,1472E-10	2,1784E-08	1
TTK	not_TF	epiFactor	D30_D0	-1,9890215	0,00029605	0,00084596	-1
TTK	not_TF	epiFactor	D46_D0	-1,831817	0,00057036	0,00145773	-1
TTYH1	not_TF	not_epi	D2_D0	-6,224167	0,00020394	0,00137545	-1
TUBA1C	not_TF	not_epi	D30_D0	-1,6257969	0,00095191	0,00231048	-1
TUBA1C	not_TF	not_epi	D46_D0	-1,6047221	0,00104594	0,00244923	-1
UBE2C	not_TF	not_epi	D30_D0	-2,2586969	5,2469E-09	1,768E-07	-1
UBE2C	not_TF	not_epi	D46_D0	-2,0851975	1,3075E-08	3,0166E-07	-1
UGP2	not_TF	not_epi	D30_D0	-1,7150758	0,00624492	0,01165168	-1
UGP2	not_TF	not_epi	D46_D0	-2,1730244	0,00129795	0,00295169	-1
VIT	not_TF	not_epi	D7_D0	4,00729039	0,00064917	0,00262794	1
VIT	not_TF	not_epi	D30_D0	4,11947496	0,00049714	0,00132147	1
VIT	not_TF	not_epi	D46_D0	6,14086583	5,6558E-06	3,1415E-05	1
WLS	not_TF	not_epi	D2_D0	7,85024007	3,5721E-13	8,9241E-10	1
WLS	not_TF	not_epi	D7_D0	8,60666583	1,4251E-13	6,2306E-10	1
WLS	not_TF	not_epi	D30_D0	7,0527393	1,0845E-12	8,2459E-10	1
WLS	not_TF	not_epi	D46_D0	6,87481963	1,421E-12	7,3087E-10	1
XG	not_TF	not_epi	D2_D0	-7,5898501	1,2858E-05	0,00015903	-1
XG	not_TF	not_epi	D7_D0	-4,0852198	9,1854E-05	0,00055907	-1
XG	not_TF	not_epi	D30_D0	-1,9315496	0,00538177	0,01021894	-1
XG	not_TF	not_epi	D46_D0	-2,7127118	0,00078936	0,00193014	-1
XRCC2	not_TF	not_epi	D30_D0	-1,4651099	0,00014247	0,00045943	-1
XRCC2	not_TF	not_epi	D46_D0	-1,1408189	0,00112019	0,00259675	-1
ZMYND10	not_TF	not_epi	D30_D0	2,82137077	0,0006472	0,0016603	1
ZMYND10	not_TF	not_epi	D46_D0	3,44893473	7,0367E-05	0,00024671	1
ZNF367	TF	not_epi	D30_D0	-1,7642869	0,00013763	0,00044548	-1
ZNF385B	TF	not_epi	D2_D0	-2,583476	0,00371159	0,01393481	-1
ZNF385B	TF	not_epi	D7_D0	2,79947811	0,00028135	0,0013517	1
ZNF552	TF	not_epi	D2_D0	-1,0147181	0,00095088	0,00462689	-1
ZYG11A	not_TF	not_epi	D7_D0	-1,5020392	0,00200702	0,00660621	-1
ZYG11A	not_TF	not_epi	D30_D0	-1,3456437	0,00422149	0,00828168	-1

Table S27. GO:BP terms for the novel tumor suppressor genes from TCGA-BRCA.

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score	Genes
1	positive regulation of neutrophil extravasation (GO:2000391)	0.001499	0.01318	999.50	6499.56	XG
2	T cell extravasation (GO:0072683)	0.001499	0.01318	999.50	6499.56	XG
3	positive regulation of Wnt signaling pathway, planar cell polarity pathway (GO:2000096)	0.001799	0.01318	799.56	5053.71	PLEKHA4
4	regulation of neutrophil extravasation (GO:2000389)	0.001799	0.01318	799.56	5053.71	XG
5	regulation of Wnt signaling pathway, planar cell polarity pathway (GO:2000095)	0.002697	0.01318	499.65	2955.69	PLEKHA4
6	regulation of postsynaptic neurotransmitter receptor activity (GO:0098962)	0.002997	0.01318	444.11	2580.41	BEGAIN
7	positive regulation of cellular extravasation (GO:0002693)	0.003296	0.01318	399.68	2284.21	XG
8	positive regulation of non-canonical Wnt signaling pathway (GO:2000052)	0.003296	0.01318	399.68	2284.21	PLEKHA4
9	positive regulation of protein localization to cell surface (GO:2000010)	0.003894	0.01385	333.03	1847.77	RIC3
10	positive regulation of myoblast differentiation (GO:0045663)	0.005388	0.01400	235.02	1227.64	BOC

Table S28. *UBE2C* positively co-expressed genes from TCGA BRCA representing altered expression across muscle differentiation

GeneID	contrast	logFC	Pvalue	Pvalue_adj	significance	D0	D2	D7	D30	D46
UBE2C	D2_D0	-0,27751	0,122964	0,220169	0	127,2681	104,9558	102,0439	26,59775	29,97714
UBE2C	D7_D0	-0,31844	0,080657	0,136839	0	127,2681	104,9558	102,0439	26,59775	29,97714
UBE2C	D30_D0	-2,2587	5,25E-09	1,77E-07	-1	127,2681	104,9558	102,0439	26,59775	29,97714
UBE2C	D46_D0	-2,0852	1,31E-08	3,02E-07	-1	127,2681	104,9558	102,0439	26,59775	29,97714
UBE2S	D2_D0	-0,85311	0,065813	0,134834	0	45,21983	24,97555	21,80253	6,548203	5,173253
UBE2S	D7_D0	-1,05286	0,02805	0,057429	0	45,21983	24,97555	21,80253	6,548203	5,173253
UBE2S	D30_D0	-2,79836	3,22E-05	0,000135	-1	45,21983	24,97555	21,80253	6,548203	5,173253
UBE2S	D46_D0	-3,12214	1,18E-05	5,72E-05	-1	45,21983	24,97555	21,80253	6,548203	5,173253
BUB1B	D2_D0	-0,96949	2,01E-05	0,000228	0	282,6046	144,3368	182,8464	38,17205	45,06637
BUB1B	D7_D0	-0,62831	0,001074	0,003966	0	282,6046	144,3368	182,8464	38,17205	45,06637
BUB1B	D30_D0	-2,88824	7,2E-11	9,75E-09	-1	282,6046	144,3368	182,8464	38,17205	45,06637
BUB1B	D46_D0	-2,64889	1,95E-10	1,56E-08	-1	282,6046	144,3368	182,8464	38,17205	45,06637
MYBL2	D2_D0	-0,41485	0,117987	0,213135	0	281,9923	211,4935	198,1347	37,55855	56,92122
MYBL2	D7_D0	-0,50908	0,06049	0,108188	0	281,9923	211,4935	198,1347	37,55855	56,92122
MYBL2	D30_D0	-2,91127	3,8E-08	7,18E-07	-1	281,9923	211,4935	198,1347	37,55855	56,92122
MYBL2	D46_D0	-2,30838	4,64E-07	4,44E-06	-1	281,9923	211,4935	198,1347	37,55855	56,92122
TROAP	D2_D0	-0,09431	0,629593	0,73697	0	45,24384	42,3512	51,86721	21,08725	19,82526
TROAP	D7_D0	0,19782	0,31778	0,415502	0	45,24384	42,3512	51,86721	21,08725	19,82526
TROAP	D30_D0	-1,09726	6,47E-05	0,000238	-1	45,24384	42,3512	51,86721	21,08725	19,82526
TROAP	D46_D0	-1,18804	2,94E-05	0,000121	-1	45,24384	42,3512	51,86721	21,08725	19,82526
CDC20	D2_D0	-0,33727	0,05257	0,113346	0	167,2348	132,3574	111,3053	27,55183	23,9415
CDC20	D7_D0	-0,58714	0,002511	0,007944	0	167,2348	132,3574	111,3053	27,55183	23,9415
CDC20	D30_D0	-2,60323	4,49E-10	3,42E-08	-1	167,2348	132,3574	111,3053	27,55183	23,9415
CDC20	D46_D0	-2,80329	1,88E-10	1,54E-08	-1	167,2348	132,3574	111,3053	27,55183	23,9415
CENPA	D2_D0	0,154498	0,422564	0,554499	0	20,60144	22,92523	28,78968	6,40351	9,575521
CENPA	D7_D0	0,482037	0,021856	0,0468	0	20,60144	22,92523	28,78968	6,40351	9,575521
CENPA	D30_D0	-1,68386	6,82E-07	6,46E-06	-1	20,60144	22,92523	28,78968	6,40351	9,575521
CENPA	D46_D0	-1,10529	5,13E-05	0,00019	-1	20,60144	22,92523	28,78968	6,40351	9,575521
KIFC1	D2_D0	-0,04003	0,785983	0,862638	0	127,21	123,6885	136,7874	39,26762	50,76878

KIFC1	D7_D0	0,105066	0,479355	0,574924	0	127,21	123,6885	136,7874	39,26762	50,76878
KIFC1	D30_D0	-1,69541	2,14E-08	4,82E-07	-1	127,21	123,6885	136,7874	39,26762	50,76878
KIFC1	D46_D0	-1,32471	3,81E-07	3,81E-06	-1	127,21	123,6885	136,7874	39,26762	50,76878
CDK1	D2_D0	0,177537	0,576369	0,692463	0	95,2995	107,897	170,7889	42,25015	59,46829
CDK1	D7_D0	0,840072	0,017566	0,039205	0	95,2995	107,897	170,7889	42,25015	59,46829
CDK1	D30_D0	-1,17499	0,002338	0,005006	-1	95,2995	107,897	170,7889	42,25015	59,46829
CDK1	D46_D0	-0,68198	0,046781	0,069596	0	95,2995	107,897	170,7889	42,25015	59,46829
KIF4A	D2_D0	0,443814	0,003413	0,013032	0	81,10089	110,3316	132,2295	31,05934	36,13971
KIF4A	D7_D0	0,705023	6,95E-05	0,000447	0	81,10089	110,3316	132,2295	31,05934	36,13971
KIF4A	D30_D0	-1,38512	4,31E-08	7,95E-07	-1	81,10089	110,3316	132,2295	31,05934	36,13971
KIF4A	D46_D0	-1,16622	3,22E-07	3,38E-06	-1	81,10089	110,3316	132,2295	31,05934	36,13971
KIF20A	D2_D0	-0,41363	0,006874	0,022809	0	154,5301	115,998	184,6793	56,31632	69,45122
KIF20A	D7_D0	0,257208	0,068276	0,119425	0	154,5301	115,998	184,6793	56,31632	69,45122
KIF20A	D30_D0	-1,45508	3,46E-08	6,73E-07	-1	154,5301	115,998	184,6793	56,31632	69,45122
KIF20A	D46_D0	-1,15367	5,28E-07	4,89E-06	-1	154,5301	115,998	184,6793	56,31632	69,45122
FOXM1	D2_D0	0,471362	0,008035	0,025779	0	139,4869	193,3581	188,272	38,84602	57,36758
FOXM1	D7_D0	0,432848	0,013195	0,030965	0	139,4869	193,3581	188,272	38,84602	57,36758
FOXM1	D30_D0	-1,84523	1,5E-08	3,76E-07	-1	139,4869	193,3581	188,272	38,84602	57,36758
FOXM1	D46_D0	-1,28147	1,02E-06	8,07E-06	-1	139,4869	193,3581	188,272	38,84602	57,36758
E2F1	D2_D0	-0,20096	0,1781	0,294247	0	55,74045	48,46654	51,57509	19,29081	24,1434
E2F1	D7_D0	-0,11146	0,444122	0,541316	0	55,74045	48,46654	51,57509	19,29081	24,1434
E2F1	D30_D0	-1,53002	5,86E-08	1E-06	-1	55,74045	48,46654	51,57509	19,29081	24,1434
E2F1	D46_D0	-1,20559	8,96E-07	7,34E-06	-1	55,74045	48,46654	51,57509	19,29081	24,1434
RAD51	D2_D0	0,016537	0,94234	0,975415	0	33,57298	33,95305	34,99069	8,006882	14,40206
RAD51	D7_D0	0,060259	0,792181	0,848669	0	33,57298	33,95305	34,99069	8,006882	14,40206
RAD51	D30_D0	-2,07744	5,7E-07	5,69E-06	-1	33,57298	33,95305	34,99069	8,006882	14,40206
RAD51	D46_D0	-1,22029	0,000116	0,000372	-1	33,57298	33,95305	34,99069	8,006882	14,40206
BRCA1	D2_D0	0,547428	0,015554	0,043718	0	52,11017	76,25142	90,03006	15,88918	34,64665
BRCA1	D7_D0	0,787233	0,001458	0,0051	0	52,11017	76,25142	90,03006	15,88918	34,64665
BRCA1	D30_D0	-1,71506	9,75E-07	8,53E-06	-1	52,11017	76,25142	90,03006	15,88918	34,64665
BRCA1	D46_D0	-0,59066	0,010521	0,018447	0	52,11017	76,25142	90,03006	15,88918	34,64665

Table S29. GO:BP terms for *UBE2C* positively co-expressed genes from TCGA BRCA

Index	Name	P-value	Adjusted p-value
1	anaphase-promoting complex-dependent catabolic process (GO:0031145)	3.365e-9	4.358e-7
2	regulation of mitotic cell cycle phase transition (GO:1901990)	2.970e-9	4.358e-7
3	regulation of cell cycle phase transition (GO:1901987)	6.286e-9	5.427e-7
4	regulation of exit from mitosis (GO:0007096)	1.659e-8	0.000001074
5	positive regulation of ubiquitin protein ligase activity (GO:1904668)	7.477e-8	0.000003402
6	mitotic spindle organization (GO:0007052)	7.880e-8	0.000003402
7	regulation of mitotic cell cycle (GO:0007346)	1.475e-7	0.000005456
8	mitotic cell cycle phase transition (GO:0044772)	3.276e-7	0.00001061
9	regulation of ubiquitin protein ligase activity (GO:1904666)	4.502e-7	0.00001296
10	positive regulation of ubiquitin-protein transferase activity (GO:0051443)	0.000001515	0.00003923

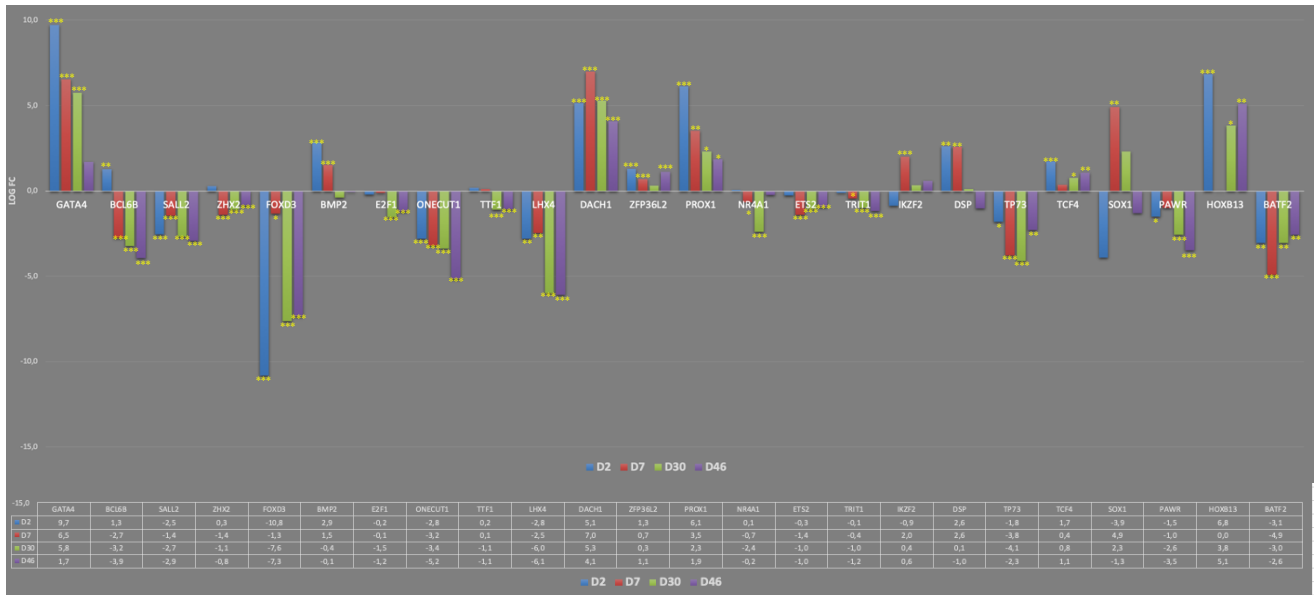


Figure S1. Involvement of cancer drivers with TFs functioning as only tumor suppressor. The comparison between muscle differentiation stages and undifferentiated cells (D0) is based on logFC and their FDR significance. *: p-value less than 0.05; **: p-value less than 0.01; ***: p-value less than 0.001.

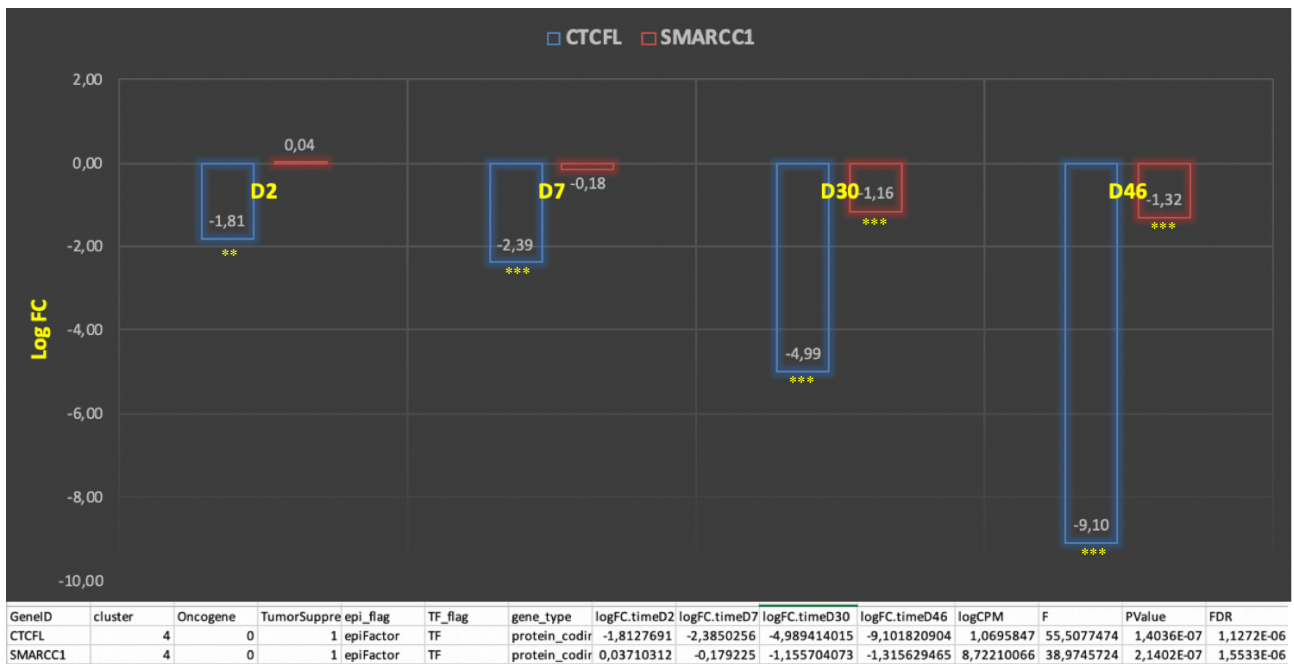


Figure S2. Altered expression of cancer drivers with roles of both TF and Epifactor functioning as tumor suppressors across muscle differentiation. The comparison between muscle differentiation stages and undifferentiated cells (D0) is based on logFC and their FDR significance. *: p-value less than 0.05; **: p-value less than 0.01; ***: p-value less than 0.001.

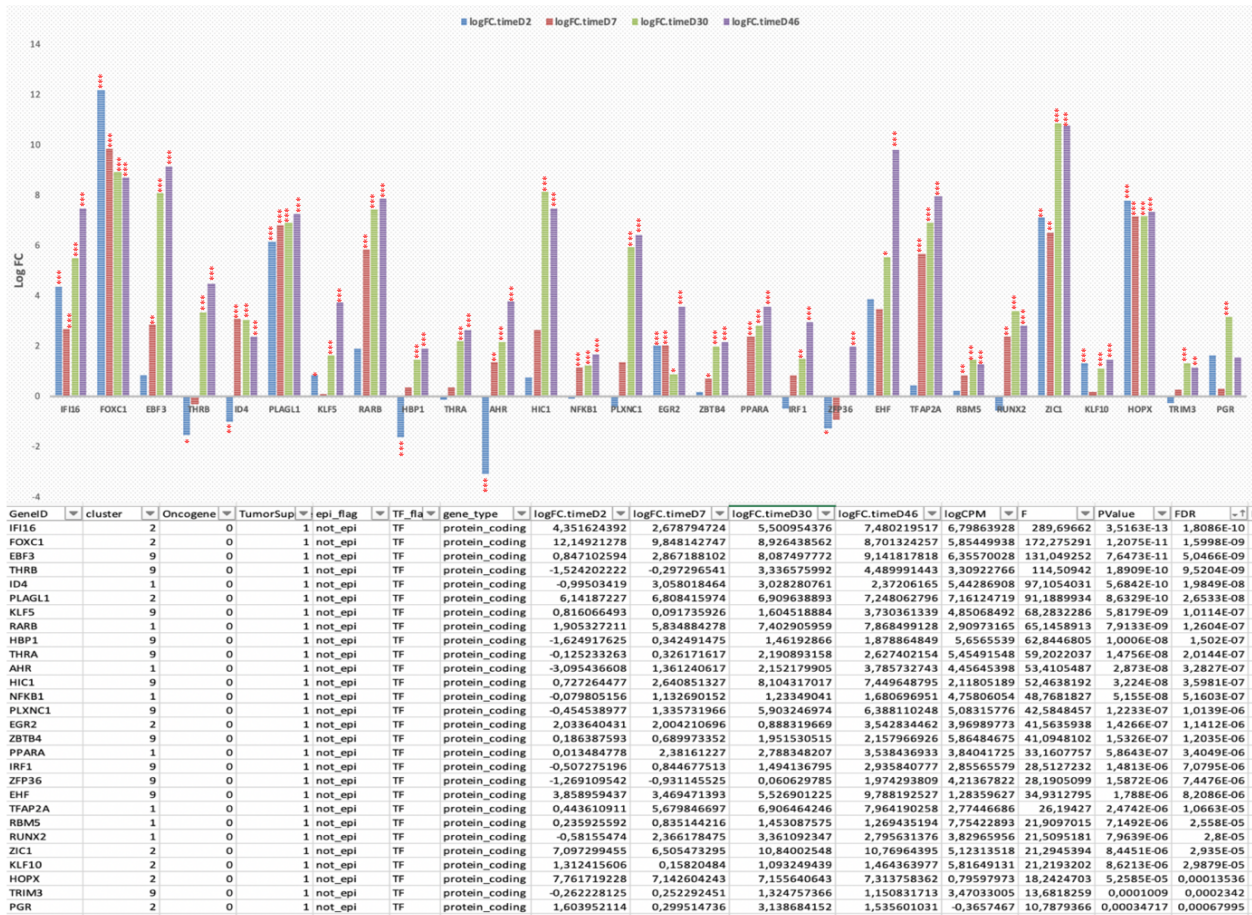


Figure S3. Involvement of TF genes with Tumor suppressor role across muscle differentiation stages. The comparison between muscle differentiation stages and undifferentiated cells (D0) is based on logFC and their FDR significance. *: p-value less than 0.05; **: p-value less than 0.01; ***: p-value less than 0.001.

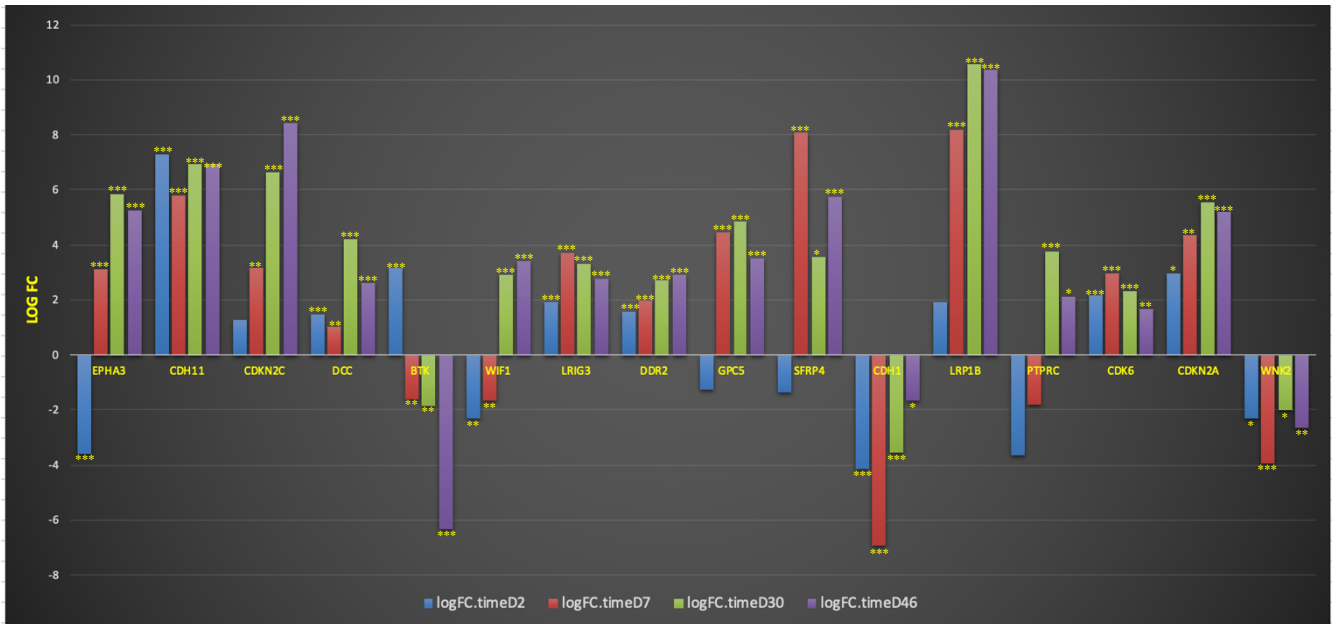
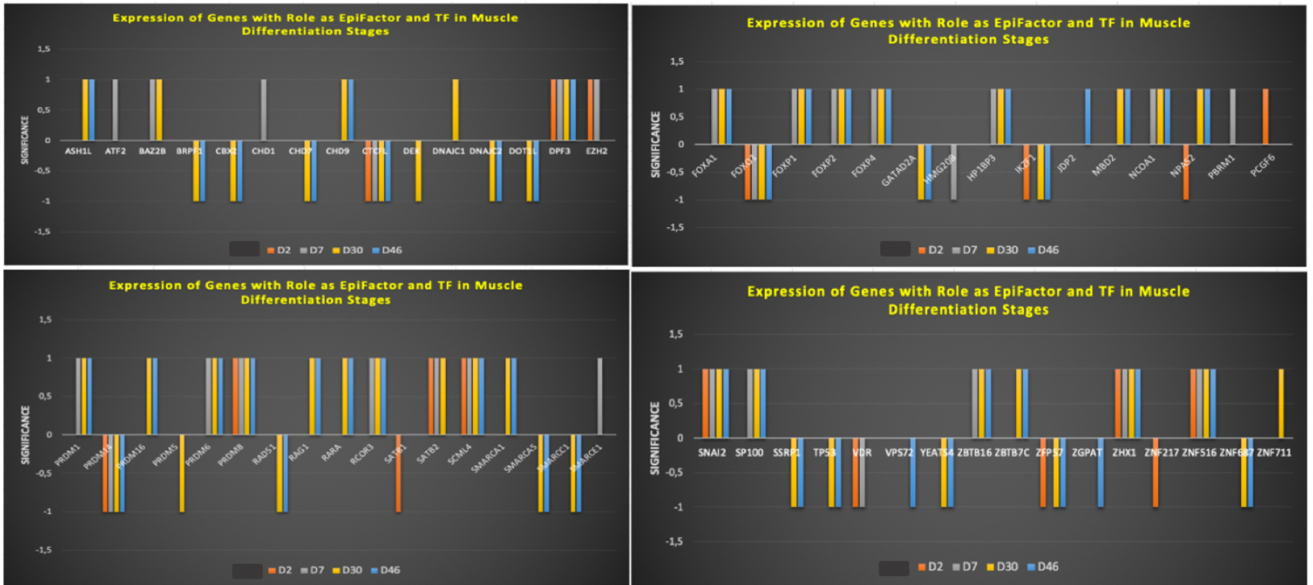


Figure S4. Altered expression of some important tumor suppressor genes and proto-oncogenes without TF and Epifactor roles across muscle differentiation. The comparison between muscle differentiation stages and undifferentiated cells (D0) is based on logFC and their FDR significance. *: p-value less than 0.05; **: p-value less than 0.01; ***: p-value less than 0.001.

A



B

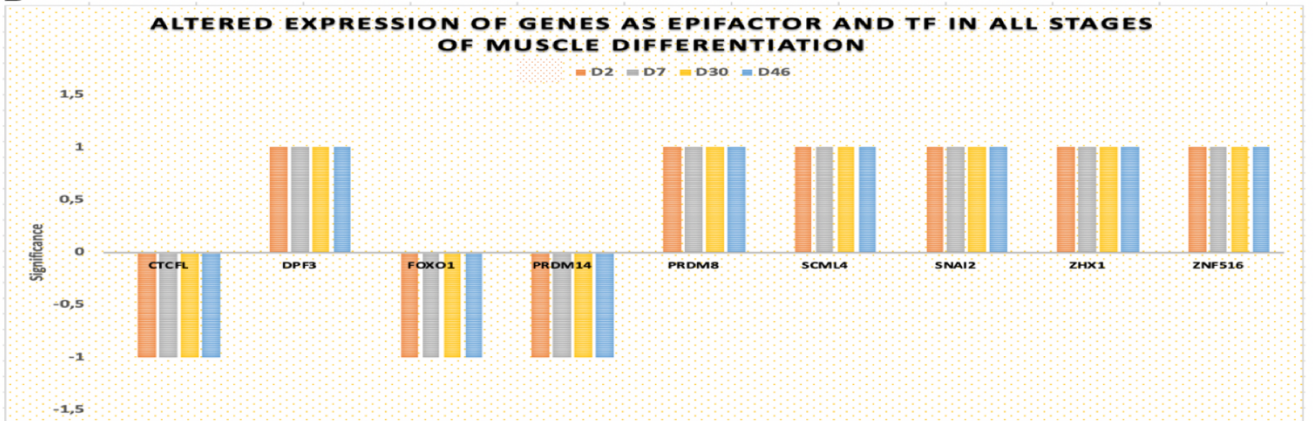


Figure S5. A. The involvement of genes with role as both TF and Epifactor in muscle differentiation stages. **B.** Genes with role as both TF and Epifactor changed at all muscle differentiation stages compared to D0.

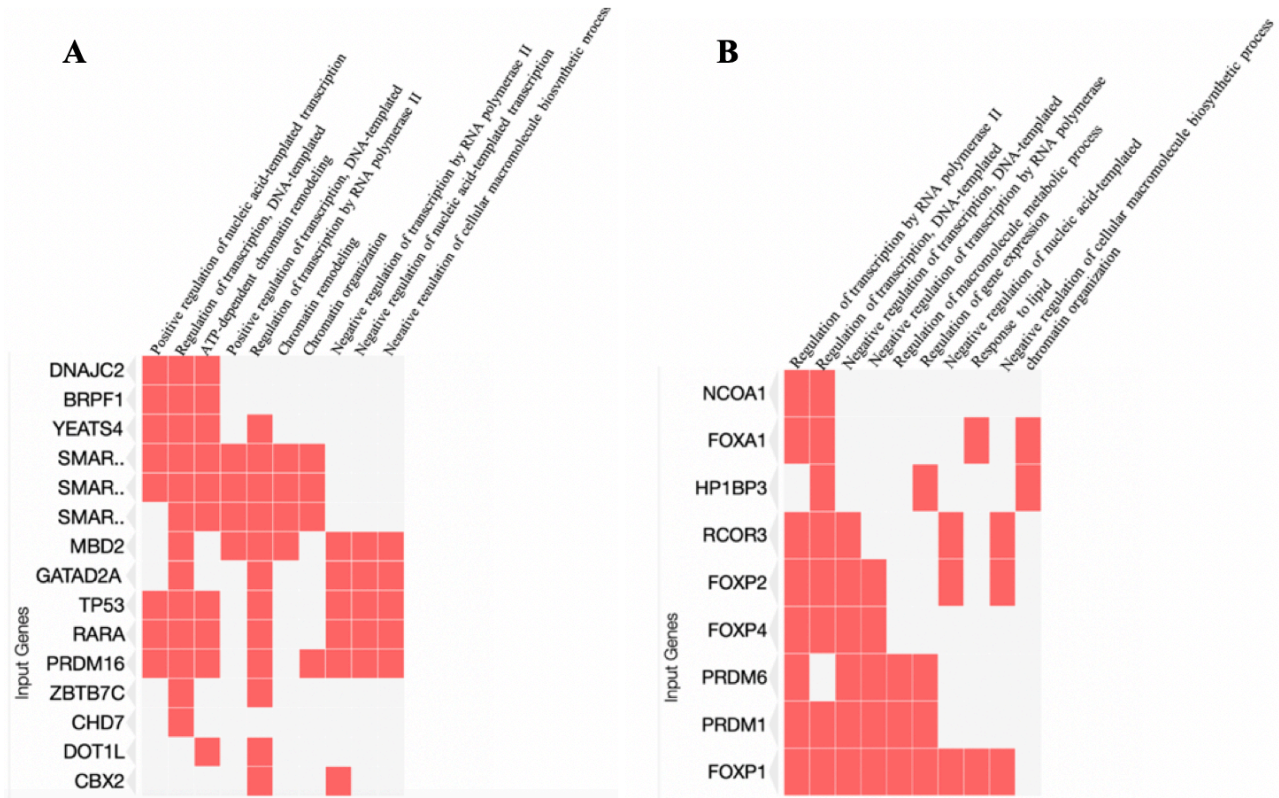


Figure S6. A. Main GO:BP terms identified for the involvement of genes with role as both TF and Epifactor with only similar altered expression in both late stages (D30 and D46 such as). **B.** Main GO:BP terms identified for the involvement of genes with role as both TF and Epifactor with only overexpression from D7 toward terminal differentiation (D30 and D46) compared to D0. GO:BP terms are listed based on adjusted p-value (the value less than 0.05 were considered statistically significant.)

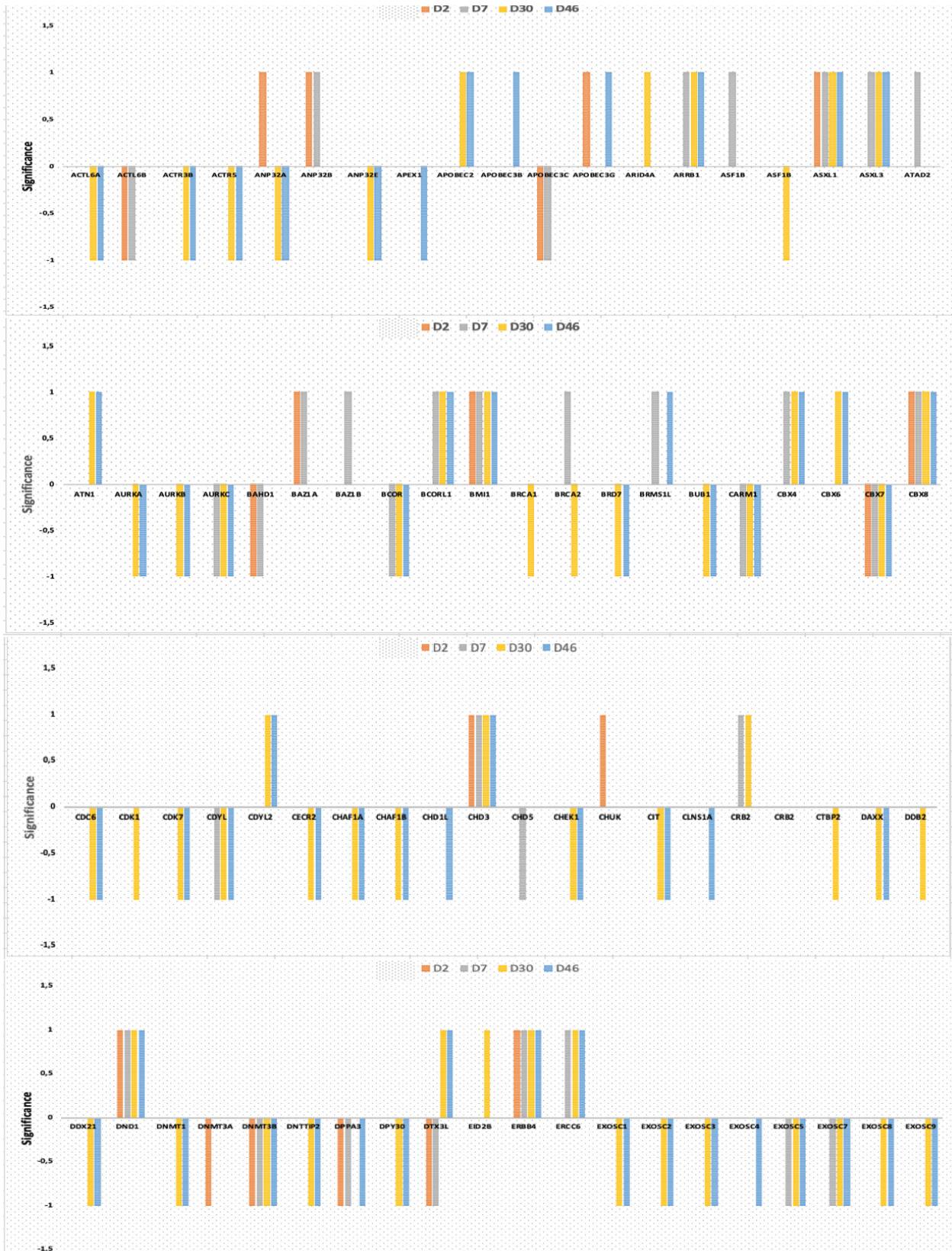


Figure S7. The involvement of genes with only Epifactor role in muscle differentiation stages. The comparison is based on significant expression value (+1 for overexpression and -1 for lower expression) compared to D0.

Figure S7 continued

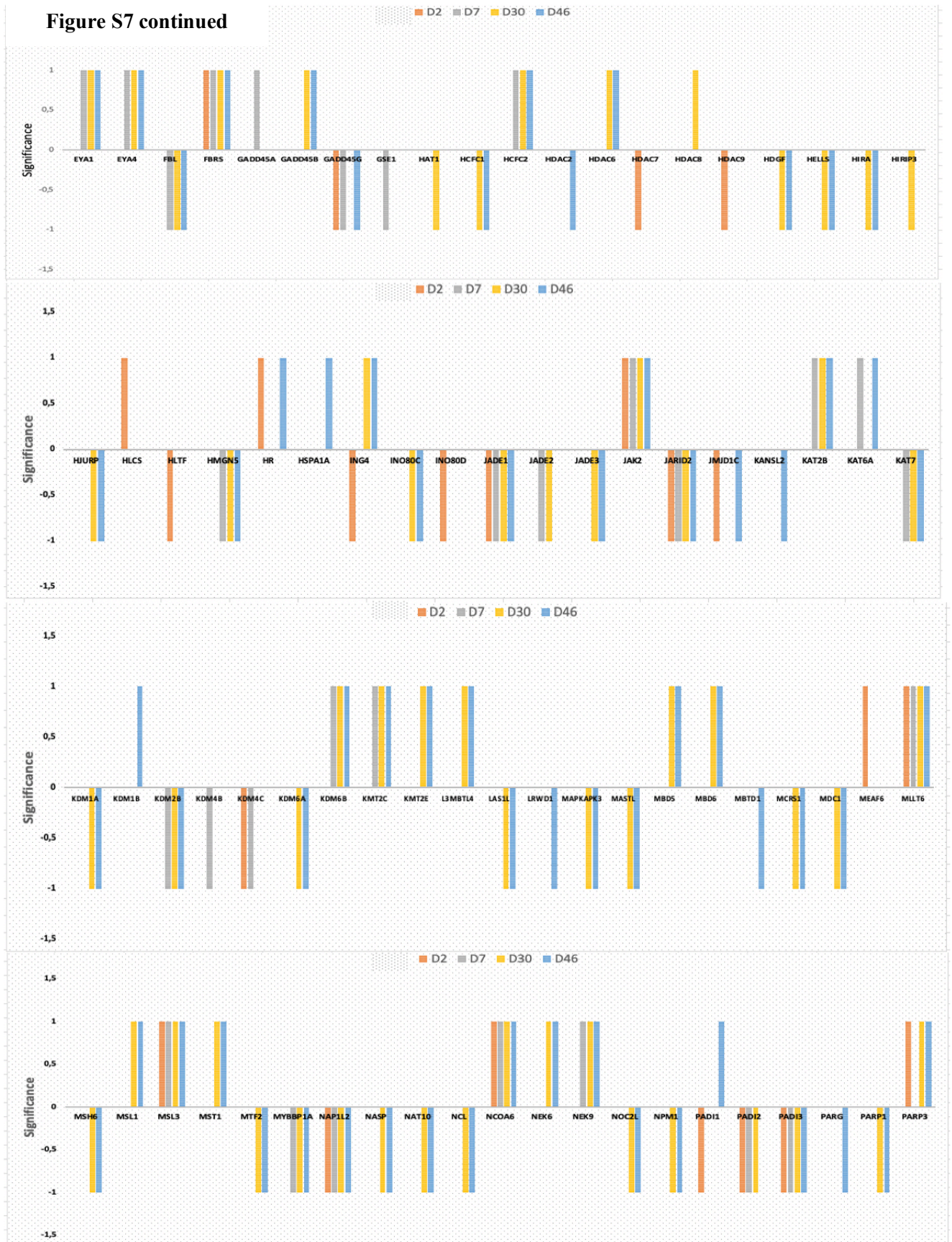
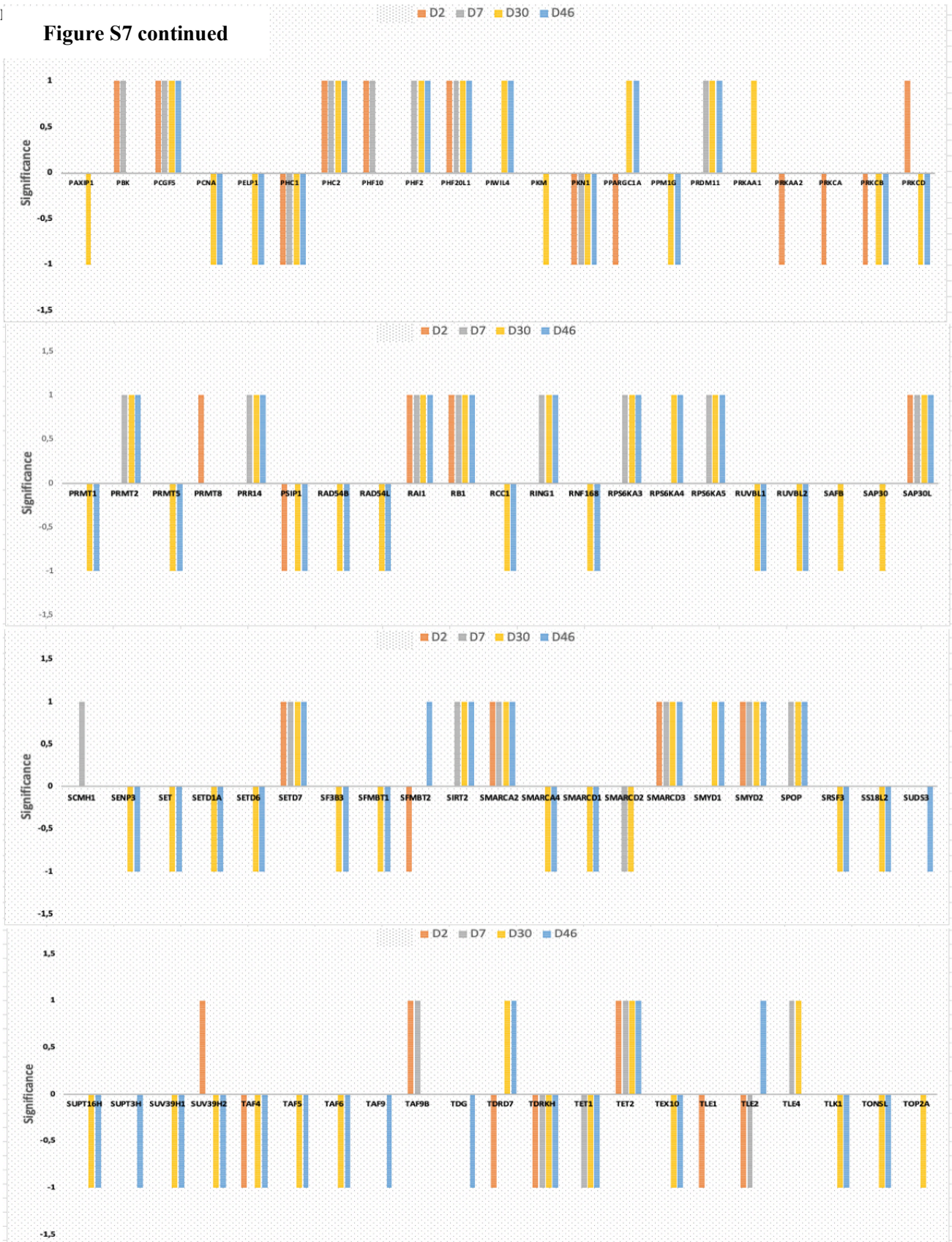


Figure S7 continued



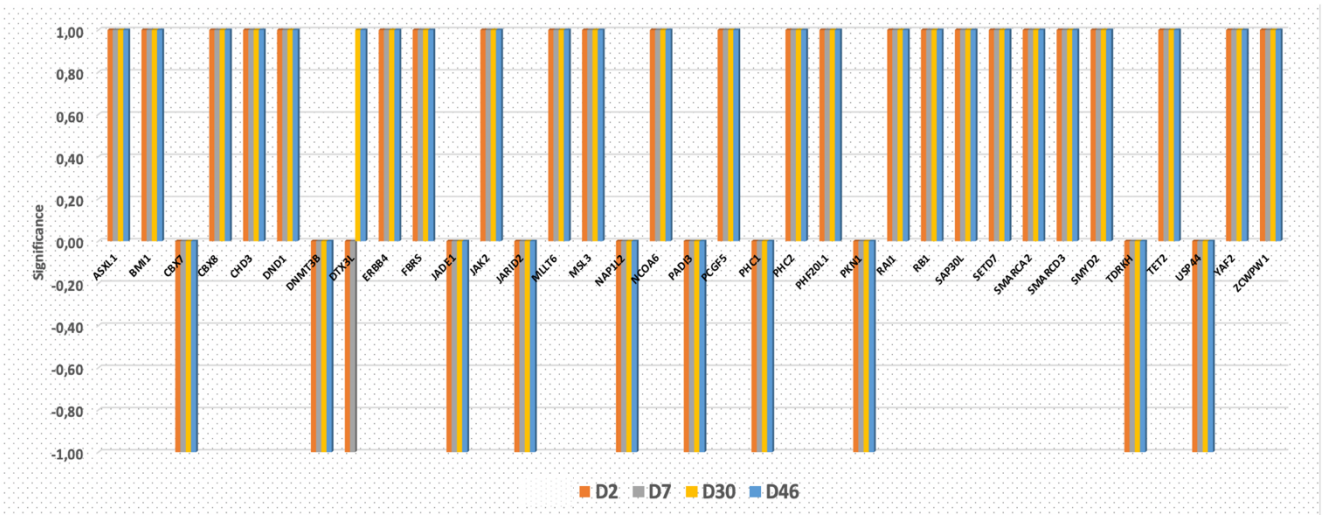
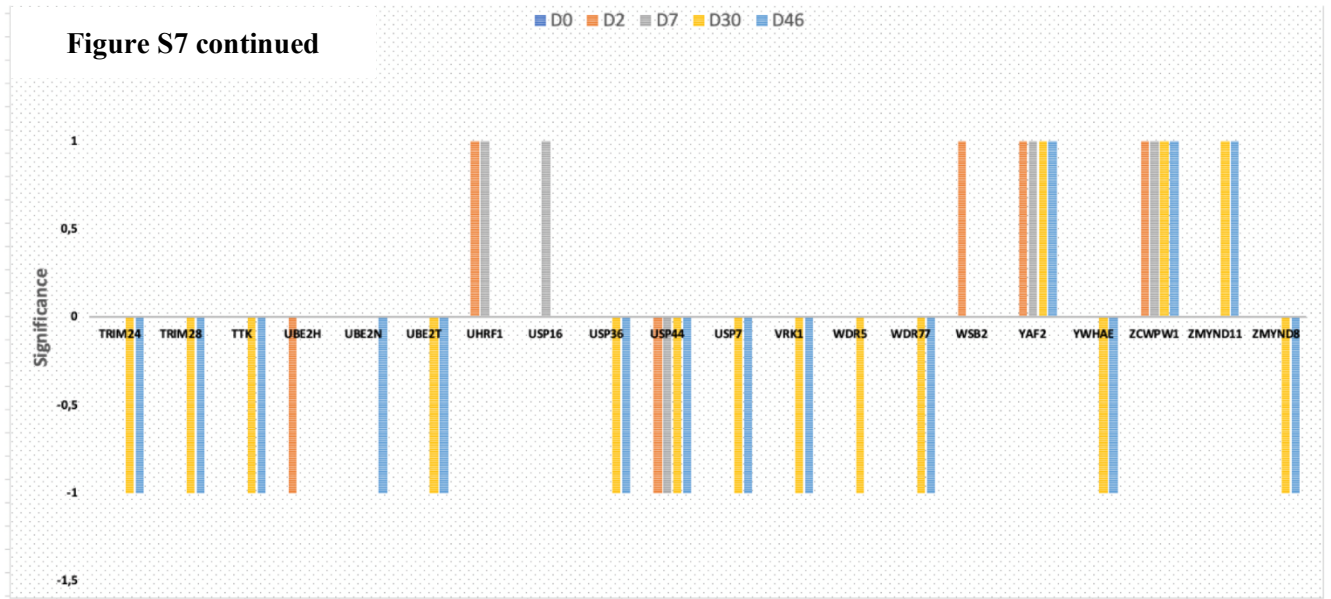


Figure S8. Altered expression of genes with only Epifactor role in all stages of muscle differentiation. The comparison is based on significant expression value (+1 for overexpression and -1 for lower expression) compared to D0.

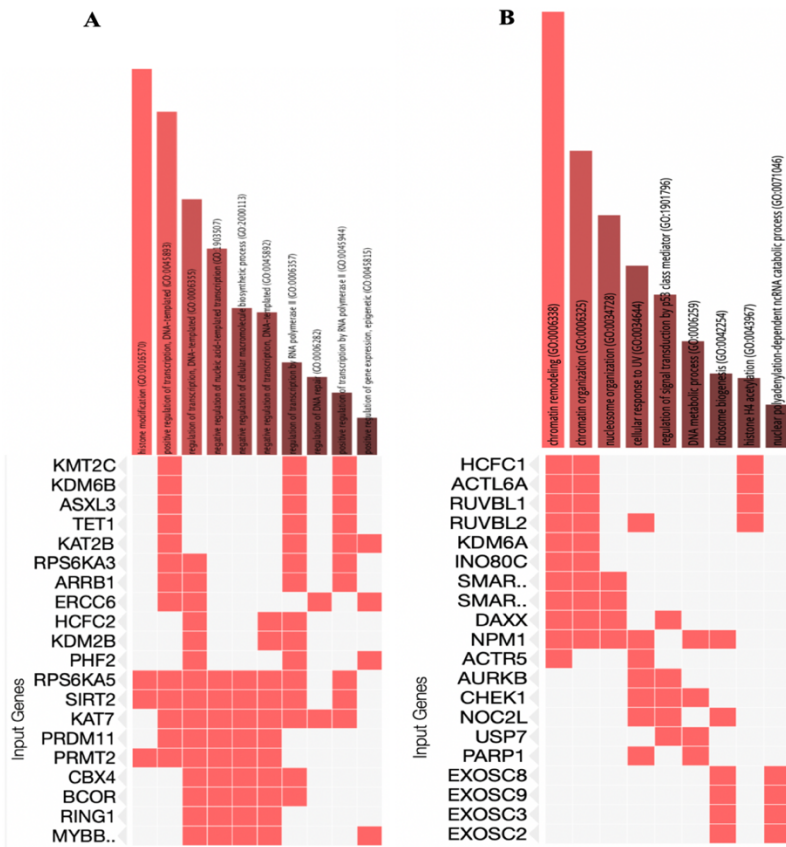


Figure S9. A. GO:BP terms of genes with expression changes from D7 toward late stages. **B.** GO:BP terms of genes with altered expression only at both D30 and D46. GO:BP terms are listed based on adjusted p-value (the value less than 0.05 were considered statistically significant.)

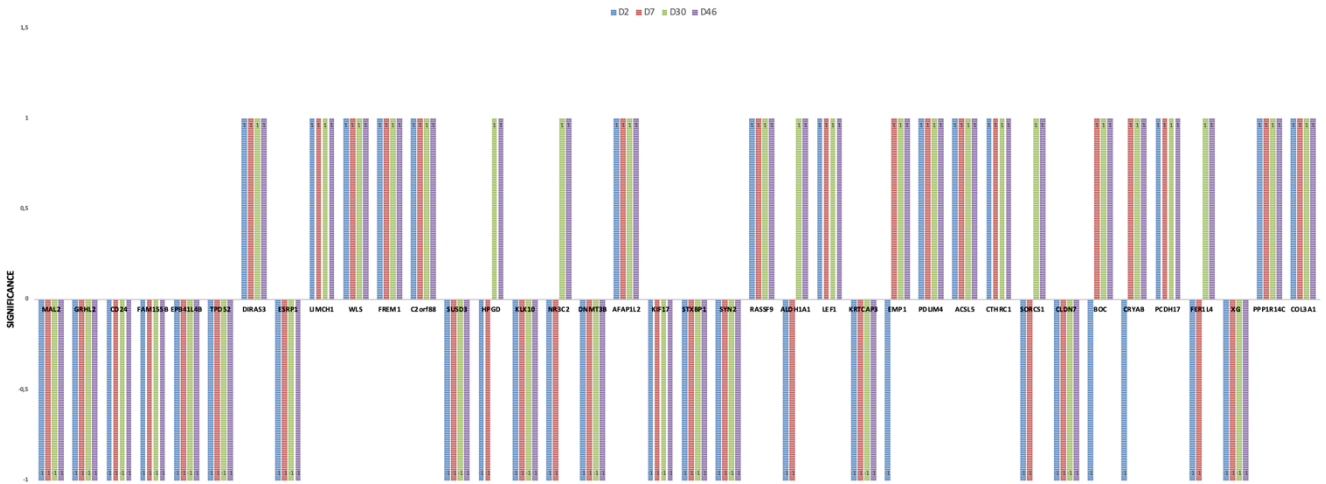


Figure S10. Altered expression of genes with significant OS from TCGA breast cancer genes in all muscle differentiation stages. The comparison is based on significant expression value (+1 for overexpression and -1 for lower expression) compared to D0.



Figure S11. Altered expression of genes with significant OS from TCGA breast cancer in specific stage (D2, D7, D30, and D46) or stages (D2 & D37, D2&D30) of muscle differentiation but not in all stages. The comparison is based on significant expression value (+1 for overexpression and -1 for lower expression) compared to D0.

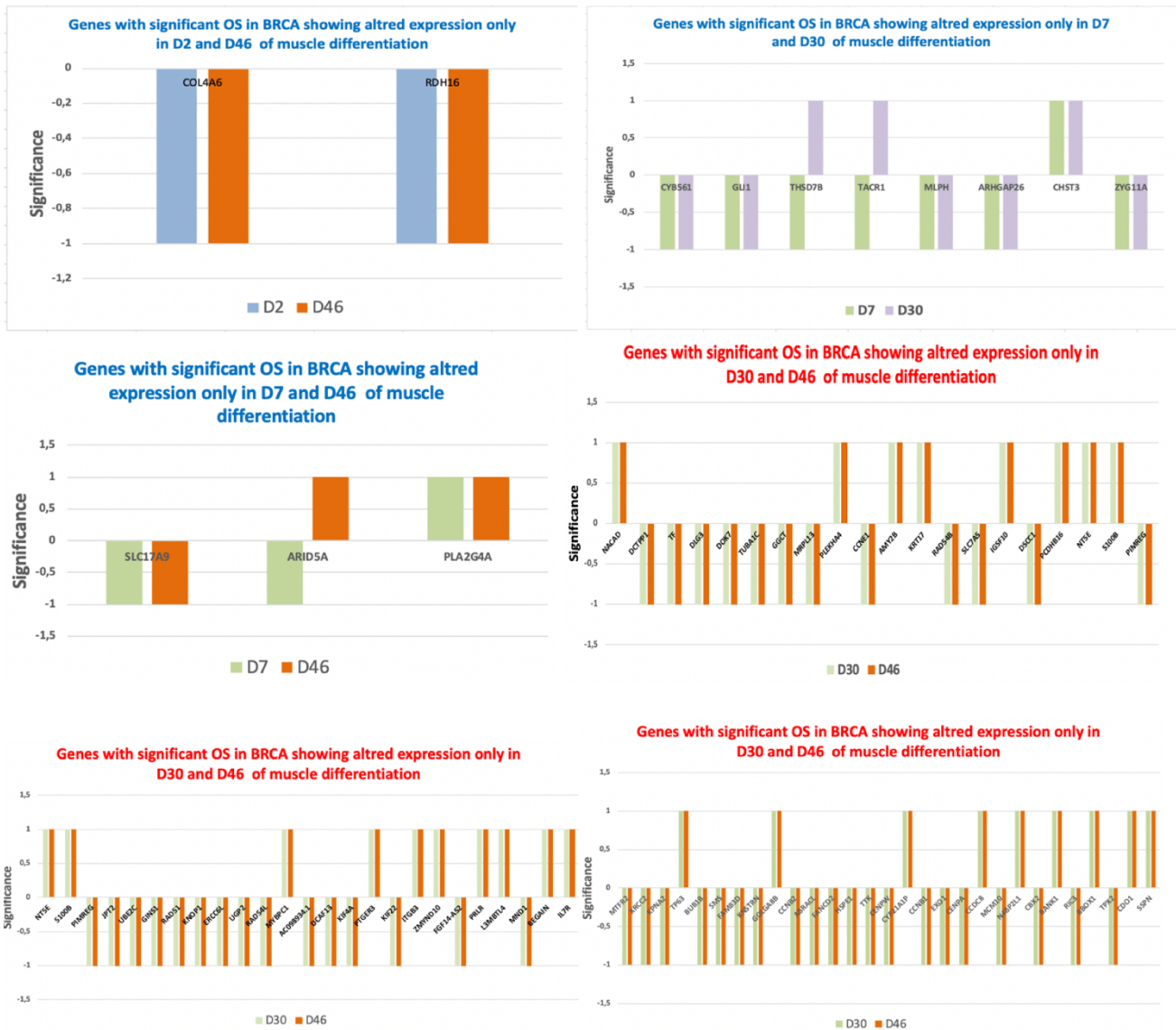


Figure S12. Altered expression of genes with significant OS from TCGA breast cancer in specific stages (D2 & D46, D7 & D30, D7 & D46, D30 & D46) of muscle differentiation but not in all stages. The comparison is based on significant expression value (+1 for overexpression and -1 for lower expression) compared to D0.

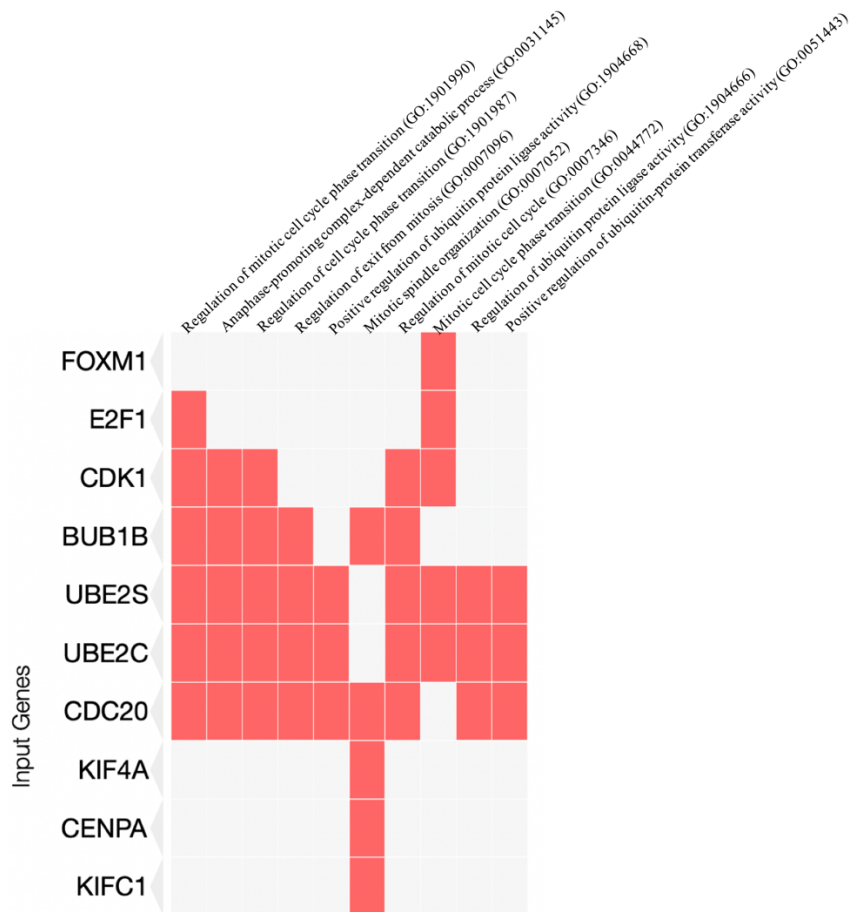


Figure S13. Enrichr GO:BP terms for *UBE2C* co-expressed genes from TCGA BRCA. GO:BP terms are listed based on adjusted p-value (the value less than 0.05 were considered statistically significant.)