Gene Reports 12 (2018) 218–224

Contents lists available at ScienceDirect

Gene Reports

journal homepage: www.elsevier.com/locate/genrep

Bioinformatics analysis of Ronin gene and their potential role in pluripotency control

Saeid Ziaei[a,b], Abdolreza Ardeshirylajimi[c], Ehsan Arefian[d], Seyed Ehsan Enderami[e], Masoud Soleimani[f], Mostafa Rezaei-Tavirani[b,*]

[a] Department of Basic Sciences, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[b] Proteomics Research Center, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[c] Department of Tissue Engineering and Applied Cell Sciences, School of Advanced Technologies in Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[d] Department of Microbiology, School of Biology, College of Science, University of Tehran, Tehran, Iran
[e] Stem Cell Technology Research Center, Tehran, Iran
[f] Department of Hematology, Faculty of Medical Science, Tarbiat Modares University, Tehran, Iran

ARTICLE INFO

Keywords:
Ronin
Reprogramming
Stem cells
Pluripotency

ABSTRACT

Pluripotent stem cells are essential cells because of their potential roles in regenerative medicine. In order to produce induced pluripotent stem cells from adult cells, we need to know more about the pluripotency genes. In this study, we analyzed the Ronin as one of the important pluripotency genes by Bioinformatics approaches. The Ronin is belonging to THAP supper family and sequence analysis shows that the THAP domain region in Ronin is highly conservative and the phylogenetic analysis shows that the Ronin is conserved from fish to mammals. Gene expression analysis shows that the expression of Ronin in differentiated tissues is negligible and there is a meaningful change during differentiation and embryo developments. In this article, we analyzed the human Ronin structure, evolutionary relations and its expression in distinct conditions. We are believed that this could be helpful to understanding the stemness procedure and could be a light for further researches.

1. Introduction

Multicellular organisms are composed of several cell types, which made up of a single cell called zygote (Smith et al., 2014). It is well known the ability of the cells to self-renew and differentiate to other multiple cell types called stemness (Clevers, 2016) (Potten and Loeffler, 1990). Embryonic stem cells (ESCs) which are derived from the inner cell mass of embryos at the blastocyst stage are immortal cells that have a huge differentiation potential for producing almost any cell other than zygote (Holland and Stanley, 2009). They are also considered as pluripotent cells that have received the most visibility because they can be converted to all types of the body's cells (Ohtsuka and Dalton, 2008). The proliferation potential of human ESCs was provided an unlimited supply for fundamental cell-based research and therapies for human diseases (Rafii and Lyden, 2003). However, there are some problems with the use of embryo-derived ESCs for cell therapy, due to ethical issues and immune response (Cowan et al., 2004; Heins et al., 2004; Hwang et al., 2004; Mateizel et al., 2005). At the beginning of the last decade, reprogrammed pluripotent cells had been recommended as an alternative to ESCs for regenerative medicine and disease modeling. It was reported that overexpression of four transcription factors including *Oct3/4, Sox2, Klf4*, and *c-Myc* can successfully convert fully differentiated somatic cells into cells with ESCs-like features, known as induced pluripotent stem cells (iPSCs) (Wilson and Wu, 2015; Yu et al., 2007; Takahashi et al., 2007). Conversion of somatic cells to the pluripotent state is a long and complex process that yields ESCs-like cells with different developmental potential. To improve the quality of iPSCs and prevent unpredictable cellular reactions – like Malignant transformation- after iPS production, the precise molecular reprogramming mechanism need to be understood (Buganim et al., 2013). Discovery of the important genes involved in the pluripotency and differentiation process is one of the most important ways to overcoming the iPS therapy problems (van den Hurk et al., 2016).

Ronin is one of the newly discovered pluripotency genes which has a distinct role in the regulation of differentiation (Baker, 2008). It indicates that Ronin expression is necessary for ESCs self-renewal and Ronin down-regulation turn cells to differentiated stage (Dejosez et al., 2008). Some research has been done on the interactions and mechanism

Abbreviations: ESCs, embryonic stem cells; iPSCs, induced pluripotent stem cells

* Corresponding author.
E-mail address: tavirani@sbmu.ac.ir (M. Rezaei-Tavirani).

https://doi.org/10.1016/j.genrep.2018.07.010
Received 23 March 2018; Received in revised form 6 July 2018; Accepted 13 July 2018

of action of Ronin in pluripotency, and it has been shown that Ronin acts as a transcriptional regulator for genes that are primarily involved in the transcription initiation, mRNA splicing, and cellular metabolism (Dejosez et al., 2010). Although other researchers have shown that Ronin is involved in the cancer progression (Parker et al., 2012; Parker, 2010), others have shown that it could acts as an anti-cancer regulator (Parker, 2010; Lian et al., 2012). Bioinformatics analysis of Ronin would be helpful for explaining the true role of Ronin in the cells. In this study, we want to analyze the Ronin gene from an evolutionary perspective by phylogenetic analysis, molecular perspective of gene structure and domain analysis and with analyzing the expression profile by looking and analyzing of some related microarray data from GEO. The function of Ronin protein in the cell will be analyzed by exploring the Ronin protein interaction network.

2. Material and methods

2.1. Data resources

The complete protein sequences of all the complete proteins named THAP11 proteins were retrieved from the NCBI Reference Sequence Database by using the Entrez search engine. (https://www.ncbi.nlm.nih.gov/refseq). Transcription data were obtained from Nucleotide Database and Expression analysis datasets were obtained from the GEO database (Edgar et al., 2002).

2.2. Domain and gene structural analysis

The functional protein domains of THAP11 were analyzed using the SMART web-based tool (Schultz et al., 2000). The gene locus and exon/intron position were identified using the Ensembl BioMart (Kinsella et al., 2011).

2.3. Phylogenetic tree construction

To create a phylogenetic tree, all complete amino acid sequences named as THAP11 in NCBI reference sequence database retrieved using the Entrez search engine. The resulting sequences were aligned using the Geneious program with default settings according to blusum62 matrix (Kearse et al., 2012). The alignment was further optimized by some manual creation using the descriptive software. The phylogenetic tree was constructed using Geneious by the UPGMA method (Sokal and Michener, 1958).

2.4. Expression analysis

Expression data was obtained from the GEO microarray database and GSE2361 and GSE18290 datasets were used respectively for human normal tissue and embryonic development expression analysis. In addition, GSE54186 was used for elucidation of Ronin changes during embryonic stem cell differentiation. Raw data from the datasets were downloaded and subsequently normalized using the R language by robust multichip average (RMA) method (López-Romero et al., 2010; Hartmann et al., 2003) and the normalized expression values were extracted and plotted.

2.5. Protein-protein interaction analysis

The protein-protein interactions of human Ronin obtained from the Biogrid database (Chatr-Aryamontri et al., 2017) and protein interaction network generated by interactions reported at least by 2 physical experiments (minimum evidence = 2) to increase the reliability of the network.

Table 1
Domain Structure Data from SMART tool, the THAP domain Region from 4 to 86 bp and a DM3 region along with the THAP from 23 to 85 bp position.

Name	Start	End	E-value
THAP	4	86	2.39e-8
DM3	23	85	1.26e-9
coiled coil	255	301	N/A

3. Results

3.1. Sequence and domain analysis

The human THAP11 protein sequence was retrieved from NCBI protein Ref sequence database and Identification of THAP11 protein domains was done by SMART search based on Multiple sequence alignments on human Thap11 protein sequence. Smart results showed that the human THAP11 consists one THAP domain and there is a DM3 region inside of it (Table 1). THAP domain as described is a DNA binding domain of the THAP supper family. DM3 region is a zinc finger region that is funded in *C. elegans* for the first time. The Biomart data showed that the human Ronin gene with 2114 base pairs (bps) transcript consists of single exon and it made a single protein with 314 residues.

3.2. Phylogenetic analysis

All the 170 sequences named as THAP11 in Refseq database were extracted and imported in Geneious software. All the resulting sequences were from vertebrata and there is not any THAP11 protein outside of this subphylum. The multiple sequence alignment showed that the most consensus parts of the THAP11 protein are THAP domain, especially DM3 region. To find more details, the THAP region of the proteins was aligned individually and it is shown that there is more than 75% conservation of all the species. The 75% conserved THAP domain in THAP11 genes was: TCCVPGCYNNSHRDKXLHFYTFPKDXELRXLWLKNXSRAGVSGCFSTFQPTTGHRXCSVHFXGGRKTYXVRVPTIFP.

The phylogenetic tree was constructed according to the multiple sequence alignment. All the Mammalia with similar THAP domains was considered as a single branch and named as “other mammalians” including human and mouse (Fig. 1). All the mammalian THAP11 have exactly the same THAP domain, except *Balaenoptera acutorostrata* (Northern minke whale), *Phascolarctos cinereus* (Koala) and *Monodelphis domestica* (gray short-tailed opossum). Lepidosauria and Archosauria are the most similar sequences to mammalians and Amphibians are in the next stage. Fishes are the most divergent group from mammalians according to THAP domain sequences. It's to be noted that there are 47% identical in THAP domain sequences between fish and mammalian THAP11 functional domain. To identify the THAP11 like proteins in outside of vertebrata, the resulting consensus sequence was aligned against the protein database with the exclusion of vertebrata taxon. The results showed that there are some proteins with 30–60% similarity with consensus THAP11 domain with unknown name or function in some insects, ants and some sea livings. The *Lingula anatine* has the closest protein to THAP11 (XP_013385818.1). *Exaiptasia pallida* have a protein named Thap11 KXJ15407.1 in and it has 44% similarity to the consensus sequence. It has 17.9% identity of the human Thap11 gene (full alignment) and 55% in THAP domain region. *Saccoglossus kowalevskii* XP_006825604 *Pseudomyrmex gracilis* XP_020297317 proteins have recorded in the NCBI protein Database as THAP domain-containing protein 11-like proteins, but the role of these proteins is not fully understood at present.