Gene 562 (2015) 40–49

Contents lists available at ScienceDirect

Gene

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

Unc-51 like kinase 1 (ULK1) in silico analysis for biomarker identification: A vital component of autophagy

Rohit Randhawa [a], Manika Sehgal [a], Tiratha Raj Singh [a], Ajay Duseja [b], Harish Changotra [a,*]

[a] Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan 1732 34 Himachal Pradesh, India
[b] Department of Hepatology, Postgraduate Institute of Medical Education and Research, Chandigarh 160 012, India

ARTICLE INFO

Article history:
Received 19 October 2014
Received in revised form 3 February 2015
Accepted 5 February 2015
Available online 19 February 2015

Keywords:
Autophagy
ULK1
Palmitoylation
Phosphorylation
Haplotype
Protein–protein interactions

ABSTRACT

Autophagy is a degradation pathway involving lysosomal machinery for degradation of damaged organelles like the endoplasmic reticulum and mitochondria into their building blocks to maintain homeostasis within the cell. ULK1, a serine/threonine kinase, is conserved across species, from yeasts to mammals, and plays a central role in autophagy pathway. It receives signals from upstream modulators such as TIP60, mTOR and AMPK and relays them to its downstream substrates like Ambra1 and ZIP kinase. The activity of this complex is regulated through protein–protein interactions and post-translational modifications. Applying in silico analysis we identified (i) conserved patterns of ULK1 that showed its evolutionary relationship between the species which were closely related in a family compared to others. (ii) A total of 23 TFBS distributed throughout ULK1 and nuclear factor (erythroid-derived) 2 (NFE2) is of utmost significance because of its high importance rate. NEF2 has already been shown experimentally to play a role in the autophagy pathway. Most of these were of zinc coordinating class and we suggest that this information could be utilized to modulate this pathway by modifying interactions of these TFs with ULK1. (iii) CATTT haplotype was prominently found with frequency 0.774 in the studied population and nsSNPs which could have harmful effect on ULK1 protein and these could further be tested. (iv) A total of 83 phosphorylation sites were identified; 26 are already known and 57 are new that include one at tyrosine residue which could further be studied for its involvement in ULK1 regulation and hence autophagy. Furthermore, 4 palmitoylation sites at positions 426, 927, 1003 and 1049 were also found which could further be studied for protein–protein interactions as well as in trafficking.

1. Introduction

Autophagy is an evolutionarily conserved degradation pathway in which cytoplasmic portions that include damaged organelles and misfolded or aggregated proteins are sequestered in double-membrane vesicles called autophagosomes. Then, these contents are delivered to the lysosomes for degradation resulting in removal/recycling of damaged/harmful contents from the cell to maintain the cellular homeostasis. This pathway is dysregulated in many diseases including neurodegenerative, inflammatory, muscle, cardiac, infectious, and neoplastic diseases. There is possibility that modulation of autophagy pathway could be helpful in better therapeutic management of these diseases.

In mammals, autophagy plays an important role in preimplantation development, survival during neonatal starvation, cell differentiation, erythropoiesis and lymphopoiesis. Autophagy is actively induced in all neonatal tissues early during development. The five identified Atg1 homologues in mammals include uncoordinated (Unc) 51-like kinase (ULK1) 1 to 4 and STK36. Carboxy-terminal domain (CTD) which is required for binding to other essential components of autophagy, Atg13 and FIP200, is lacking in ULK3, ULK4 and STK36. Therefore, ULK1 and ULK2 are the primary candidate mammalian Atg1 orthologues, essential for induction of autophagy. ULK1 protein expression pattern studies have been done (Kundu et al., 2008). ULK1 knock out mouse model was viable and did not show any evident developmental defects, in contrast to other core autophagy genes (Atg3, Atg5, Atg7, Atg9 and Atg16L1) where their deletions led to neonatal lethality. ULK1 expression levels were elevated during erythroid maturation but not of ULK2 suggesting that ULK2 was not involved in this process. Moreover, they also showed an important role of ULK1 in selective clearance of mitochondria and ribosomes in reticulocytes. The reasons that ULK1 is not essential for murine survival could be (i) ULK2, which shows >50% homology with ULK1 and shows functional redundancy and induce autophagy and/or (ii) existence of ULK1 independent mechanism of autophagy. Furthermore, Chan and co-workers have shown that in HEK293 cells ULK1 was critical for inducing autophagy in response to amino acid starvation (Chan et al., 2009). Therefore, the focus of this study was to analyze ULK1 which is a major regulator of autophagy.

ULK1, a serine–threonine kinase, is one of the central human autophagy-related genes and its chromosomal location is 12q24.3. A

* Corresponding author.
E-mail addresses: hchangotra@yahoo.com, harish.changotra@juit.ac.in (H. Changotra).

http://dx.doi.org/10.1016/j.gene.2015.02.056
0378-1119/

ULK1 gene is 28,517 bp long with 28 exons and is translated to 1050 amino acids. ULK1 forms a stable complex with Atg13, FIP200, and Atg101. This complex plays a crucial role in initiation step of autophagy. ULK1 regulates its substrates and is itself regulated by phosphorylation events. mTOR1, AMPK and TIP60 are its well known upstream regulators. It is hyperphosphorylated in nutrient-rich conditions and dephosphorylates on starvation. So far, around 30 phosphorylation sites have been identified on ULK1 and most of the kinases responsible for its phosphorylation and functions of these are still unidentified (Mack et al., 2012). This supports that phosphorylation events play an important role in ULK1 regulation. Recently, decreased expression of ULK1 has been shown in breast cancer patients, which was associated with cancer progression and low autophagic activity (Tang et al., 2012). However, another study showed higher expression of ULK1 in the hepatocellular carcinoma (HCC) patients and furthermore, higher ULK1 expression in HCC patients was associated with low survival rate (Xu et al., 2013). These studies indicate different roles of autophagy in different types of cancers and indeed in different diseases. Therefore, ULK1 could be used as a prognostic marker for cancer patients. Moreover, this gene has been shown to be involved in genetic susceptibility of Crohn's disease (CD). Recent studies have shown the association of three SNPs (rs12303764, rs10902469 and rs7488085) with CD (Henckaerts et al., 2011). These variations could be used as prognostic markers in the therapeutic interventions after validation in more number of patients and other populations.

In the present study, we computationally analyzed the ULK1 gene for its phylogeny reconstruction which suggests that it is closely related in a family. We identified new TFBS, snSNP with their possible phenotypic effect on ULK1 protein function, phosphorylation and palmitoylation sites and protein–protein interactions. This comprehensive in silico analyses would be helpful to unravel the functions of this gene and understand autophagy as well as non-autophagy roles of this gene.

2. Material and methods

In the study, an extensive examination of the ULK1 gene is carried out which is subdivided into seven major sections; (1) Identification of site-specific residues and phylogenetic analysis; (2) Regulatory elements and over-represented transcription factor binding site (TFBS) recognition; (3) Detection of nsSNPs, their phenotypic effects and quantitative statistical analysis for genetic parameters; (4) Elucidation of putative phosphorylation and palmitoylation sites; and (5) Protein–Protein Interaction (PPI) studies.

2.1. Identification of site-specific residues and phylogenetic analysis for ULK1

The analyses initiated with retrieval of human protein sequence for ULK1 (GenBank Accession Number: AAC32326) from the National Center for Biotechnology Information (NCBI) and corresponding protein sequences for other seven species of families Hominidae (*Pan troglodytes*; GenBank Accession Number: JAA43195), Bovidae (*Bos taurus*; GenBank Accession Number: NP_001192856), Cricetidae (*Cricetulus griseus*; GenBank Accession Number: EGW02429), Pteropodidae (*Pteropus alecto*; GenBank Accession Number: ELK14239), Muridae (*Rattus norvegicus*; GenBank Accession Number: NP_001101811, *Mus musculus*; GenBank Accession Number: NP_033495), Pipidae (*Xenopus* (*Silurana*) *tropicalis*; GenBank Accession Number: NP_001106388) were also retrieved and further deliberated for their evolutionary conservation.

2.1.1. Evolutionary conserved and variable regions

The genetic variations leading to different phenotypes were analyzed by observing the variable regions in the multiple sequence alignment (MSA) generated for the ULK1 gene. The latter was carried out using multiple sequence comparison by log-expectation (MUSCLE) (Edgar, 2004) and multiple alignment using fast Fourier transform (MAFFT) (Katoh et al., 2002). These programs use log-expectation scores and fast Fourier transform methods respectively for providing better average accuracy and speed compared to other MSA algorithms. Programs were used with their default parameters.

2.1.2. Evolutionary relationship associated with ULK1

Highly conserved regions play an imperative role in phylogenetic tree reconstruction. Therefore, the evolutionary relationship among eight species was elucidated on the basis of sequence similarities by

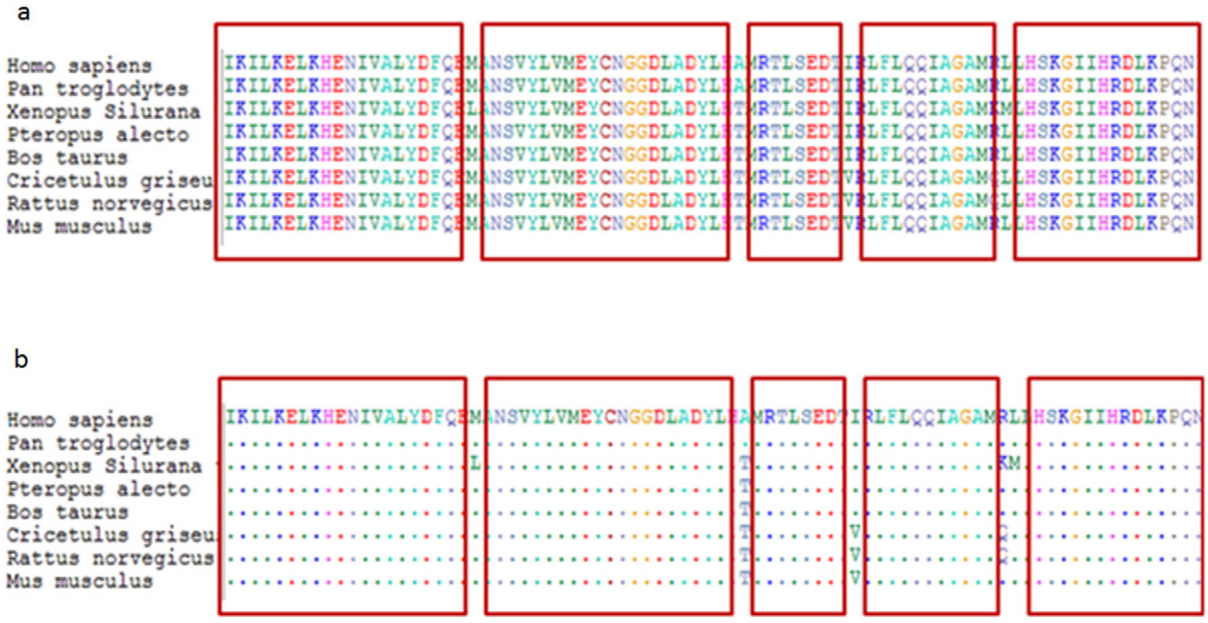

Fig. 1. Partial representation of Multiple Sequence Alignment of ULK1 gene for 8 different species. This was carried out using multiple sequence comparison by log-expectation (MUSCLE) and multiple alignment using fast Fourier transform (MAFFT) which use log-expectation scores and fast Fourier transform methods, respectively. Programs were used with their default parameters. Human sequence was taken as a reference and is shown at the top of MSA. Areas in boxes represent various conserved regions in MSA.