Available online at www.sciencedirect.com
SCIENCE DIRECT
Gene 346 (2005) 97–104
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

The 350-fold compacted Fugu parkin gene is structurally and functionally similar to human *Parkin*

Wei-Ping Yu[a,b], Jeanne M.M. Tan[c], Katherine C.M. Chew[c,d], Tania Oh[e], Prasanna Kolatkar[e], Byrappa Venkatesh[b], Ted M. Dawson[f], Kah Leong Lim[c,d,*]

[a]*Gene Regulation Laboratory, National Neuroscience Institute, 11 Jalan Tan Tock Seng, Singapore 308433, Singapore*
[b]*Institute of Molecular and Cell Biology, 61 Biopolis Drive, Singapore 138673, Singapore*
[c]*Neurodegeneration Research Laboratory, National Neuroscience Institute, 11 Jalan Tan Tock Seng, Singapore 308433, Singapore*
[d]*Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117542, Singapore*
[e]*Genome Institute of Singapore, 60 Biopolis Street, Singapore 138672, Singapore*
[f]*Institute for Cell Engineering, Departments of Neurology and Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD 21205, United States*

Received 18 June 2004; received in revised form 9 September 2004; accepted 28 September 2004
Available online 15 December 2004
Received by A.J. van Wijnen

Abstract

Mutations in the human parkin gene (*huParkin*) are the predominant genetic cause of familial parkinsonism. The *huParkin* locus, spanning about 1.4 Mb, is one of the largest in the human genome. Despite its huge size, *huParkin* codes for a rather short transcript of about 4.5 kb. To gain an insight into the structure, function and evolutionary history of *huParkin*, we have characterized the pufferfish [*Fugu rubripes* (Fugu)] ortholog of *huParkin*. A remarkable feature of the Fugu parkin gene (*fuparkin*) is its unusually compact size. It spans only about 4 kb and is thus 350-fold smaller than its human ortholog. The Fugu and human parkin genes are otherwise highly similar in their genomic organization and expression pattern. Furthermore, like human Parkin, Fugu parkin also functions as an ubiquitin ligase. These shared features between *fuparkin* and *huParkin* suggest that the physiological function and regulation of the parkin gene are conserved during the evolution of vertebrates. Conceivably, the compact locus of *fuparkin* could serve as a useful model to understand the transcriptional regulation of *huParkin*.

Keywords: Parkinson's disease; Ubiquitin ligase; Transcriptional regulation; Dystrophin

1. Introduction

Parkinson's disease (PD) is a major neurodegenerative movement disorder affecting approximately 1% of the elderly population worldwide (Siderowf and Stern, 2003). Although the etiology of PD remains poorly understood, the identity of several key genetic players whose mutations cause rare cases of familial parkinsonism has recently been elucidated (Lim and Lim, 2003). Of these, mutations in the parkin gene are currently recognized as the main contributor to familial parkinsonism (Kitada et al., 1998; Lucking et al., 2000). Emerging evidence also suggests a role for parkin in idiopathic PD and that parkin haploinsufficiency may be sufficient for the disease (West et al., 2002a; Foroud et al., 2003). To date, a wide range of parkin mutations, including several missense/nonsense/frameshift point mutations as well as exon deletion, duplication and triplication have

Abbreviations: PD, Parkinson's disease; RING, really interesting new gene; *PACRG*, parkin coregulated gene; ORF, open reading frame; PCR, polymerase chain reaction; RT, reverse transcription; RACE, rapid amplification of cDNA ends; TSS, transcriptional start site.

* Corresponding author. Neurodegeneration Research Laboratory, National Neuroscience Institute, 11 Jalan Tan Tock Seng, Singapore 308433, Singapore. Tel.: +65 6357 7520; fax: +65 6256 9178.
E-mail address: Kah_Leong_Lim@nni.com.sg (K.L. Lim).

0378-1119/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2004.09.022

been identified in parkin-related PD patients (Mata et al., 2004). It is currently unknown whether the heterogeneity in parkin mutations may in part be responsible for a wide age of onset (7–58 years) observed in patients with parkin-associated parkinsonism.

The human parkin gene (*huParkin*), located on chromosome 6q26 (#6:161, 740, 081-163, 119, 211 May 04 Golden Path assembly), encodes a 465-amino-acid protein comprising a ubiquitin-like (UBL) domain at its N-terminus, a really interesting new gene (RING) box domain at its C-terminus and a unique middle segment that links the two domains (Kitada et al., 1998; Lim and Lim, 2003). The RING box domain of Parkin can be further subdivided into three consecutive cysteine-rich protein domains, namely, two C_3HC_4 RING fingers (RING 1 and 2) with a peculiar C_6HC domain (IBR) separating them (Morett and Bork, 1999). At the extreme C-terminal end of the protein is a class II PDZ-binding motif bearing the amino acid sequence FDV that is important for the subcellular localization of Parkin within postsynaptic densities in the brain (Fallon et al., 2002). The consensus residues of the RING box domain and PDZ motif are identical in humans, rat and mouse (Gu et al., 2000; Fallon et al., 2002). Several groups have demonstrated that Parkin functions as an ubiquitin ligase associated with intracellular protein homeostasis and that the RING box domain is essential for its catalytic activity (Imai et al., 2000; Shimura et al., 2000; Zhang et al., 2000). Notably, many clinically relevant point mutations on Parkin occur on the RING box region.

The *huParkin* locus, spanning about 1.4 Mb, is one of the largest in the human genome. It contains 12 exons and codes for a 4.5-kb transcript (Kitada et al., 1998; West et al., 2003). The largest known human gene to date is *Dystrophin*, which spans about 2 Mb. However, in contrast to *huParkin*, *Dystrophin* is composed of 79 exons and codes for a 13-kb transcript (Chelly et al., 1988; Ahn and Kunkel, 1993). Whereas the coding sequences of *Dystrophin* represents about 0.5% of the total length of the gene (Ahn and Kunkel, 1993; Pozzoli et al., 2003), the coding sequences of *huParkin* occupy only 0.1% of the gene length. It thus appears that *huParkin* has one of the highest ratio of noncoding to coding DNA lengths in the human genome. The role of the large introns which account for about 99.9% of *huParkin* is not known. Interestingly, the promoter region of *huParkin* is very short. Its 5′ flanking gene, parkin coregulated gene (*PACRG*), resides just 204-bp upstream on the opposite strand (Asakawa et al., 2001; West et al., 2003). Given the close head-to-head linkages of the two genes, it is not surprising that their promoters overlap (West et al., 2003). It is possible that the two genes share some regulatory elements. However, despite its key role in familial parkinsonism, very little is known about the transcriptional regulation of *huParkin*.

To gain an insight into the structure, function and regulation of *huParkin*, we have characterized the parkin gene from the pufferfish, *Fugu rubripes* (Fugu). At 400 Mb, Fugu has the smallest vertebrate genome but codes for a similar gene repertoire to human (Venkatesh et al., 2000; Aparicio et al., 2002). The compact size of the Fugu genome is therefore useful in gene discovery as well as in understanding vertebrate gene structure and regulation. Typically, Fugu genes are about eightfold smaller compared to the human genes, in proportion to their genome sizes. However, we found that the size of the Fugu parkin gene (*fuparkin)* is dramatically compressed by about 350-fold compared to its human ortholog. Despite the huge reduction in the size, the structure, expression and biochemical function of *fuparkin* are similar to its human ortholog. The conserved features between the Fugu and human parkin genes suggest that the two genes may be similarly regulated.

2. Materials and methods

2.1. Database search and in silico analysis

Sequences for the parkin gene from human, rat, mouse, *Anopheles*, *Drosophila* and *C. elegans* were retrieved from the UCSC Human Genome Browser at http://genome.ucsc.edu or the NCBI database at http://www.ncbi.nlm.nih.gov. As the parkin gene is only partially annotated in some of these species, we annotated the exons by iterative comparison of the genomic sequence with the cDNA sequence. To identify the genomic sequence of *fuparkin*, we searched the draft Fugu genome sequence at http://www.fugu-sg.org using the human Parkin protein sequence as the query. The genomic structure of *fuparkin* and its flanking genes were predicted using GENSCAN (http://genes.mit.edu/GENSCAN.html). The exon–intron organization of *fuparkin* was confirmed by sequencing the cDNA amplified by reverse transcription (RT)–polymerase chain reaction (PCR) (see below).

2.2. 5′ RACE, Cloning and RT–PCR analysis of Fugu parkin transcripts

The cDNA templates used in these experiments were reverse-transcribed from the total RNAs of various Fugu tissues using the SMART rapid amplification of cDNA ends (RACE) cDNA Amplification kit (Clontech, USA). We used the 5′ RACE approach to determine the transcription start site and the first exon of *fuparkin*. A nested PCR amplification of the cDNAs from Fugu brain was performed according to the manufacturer's instruction. The two gene-specific reverse primers used, 5′-GCAGAAGATCTGGCTGAACTCCCTGC-3′ and 5′-CAGTGGCCTCTTCCTGCAGCTCTACG-3′, were derived from the putative second exon of *fuparkin*. The resulting PCR product was cloned into pGEM-T vector (Promega, USA) and sequenced completely. The full-length cDNA of Fugu parkin was amplified by PCR using the primers 5′-ACGAAGGTC-