Available online at www.sciencedirect.com
ScienceDirect
Virology 362 (2007) 1 – 5
VIROLOGY
www.elsevier.com/locate/yviro

Rapid Communication

Insertion of an HERV(K) LTR in the intron of *NBPF3* is not required for its transcriptional activity

Karl Vandepoele, Frans van Roy *

Department for Molecular Biomedical Research, VIB, B-9052 Ghent, Belgium
Department of Molecular Biology, Ghent University, B-9052 Ghent, Belgium

Received 7 December 2006; returned to author for revision 8 January 2007; accepted 21 January 2007
Available online 28 March 2007

Abstract

The *NBPF* genes are members of a recently described gene family that has an intricate genomic organization. These genes can be subdivided into two subfamilies based on the presence of an intronic HERV(K) LTR insertion in the 5′ region. A recent report describes a functional implication for this insertion, claiming that only *NBPF* genes with this insertion are transcriptionally active [Illarionova, A.E., Vinogradova, T.V., Sverdlov, E.D., 2007. Only those genes of the KIAA1245 gene subfamily that contain HERV(K) LTRs in their introns are transcriptionally active. Virology, 358 (1): 39–47]. Here, we show that an *NBPF* gene lacking this insertion, *NBPF3*, is expressed in a variety of tissues. Thus the effect of HERV(K) LTR insertion on *NBPF* gene expression remains unknown.

Keywords: Human endogenous retrovirus; LTR; Gene families; Regulation of gene expression; Human genome; Evolution

Introduction

The *NBPF* gene family consists of about 22 genes with high sequence similarity. They are located primarily on regions of segmental duplications on chromosome 1 and have undergone recent partial and complete gene duplications (Vandepoele et al., 2005). Additionally, several reports have shown that the copy number of these genes is variable in humans (Redon et al., 2006; Tuzun et al., 2005). The proteins encoded by these genes are composed of repetitive elements for which no functional role has been described so far. Recently, two subfamilies were identified among the *NBPF* genes, only one of which has an intronic HERV(K) LTR insertion in the 5′ region (Illarionova et al., 2007). This HERV(K) LTR sequence is derived from the long terminal repeat of a human endogenous retrovirus of the K family and contains a set of regulatory elements necessary for transcription of the retroviral RNA (Sverdlov, 1998). It is estimated that 5.3–8.3% of the human genome consists of these kinds of transposable elements (IHGSC, 2001; Venter et al., 2001). Insertion of these elements near a gene can affect its expression (Buzdin et al., 2006). It was reported that insertion of this LTR is involved in transcriptional activation of *NBPF* genes, and that genes lacking this insertion are not expressed (Illarionova et al., 2007). Here, we show that *NBPF3*, located on chromosome 1p36, lacks this insertion but is nevertheless expressed in a variety of tissues.

Results

In a study of the *NBPF* genes, Illarionova et al. (2007) considered the cDNA clone KIAA1245, a *NBPF11* transcript, as a prototypical *NBPF* transcript. Upon comparing the KIAA1245 clone to genomic sequences, these authors identified a number of related sequences corresponding to different members of the recently described *NBPF* family (Vandepoele et al., 2005). An HERV(K) LTR insertion was identified in an intron in the 5′ region of a number of these paralogs, and it was claimed that only the *NBPF* genes with this insertion are transcriptionally active (Illarionova et al., 2007).

* Corresponding author. Department for Molecular Biomedical Research, VIB-Ghent University, Technologiepark 927, B-9052 Zwijnaarde, Belgium. Fax: +32 9 33 13 609.
E-mail address: Frans.VanRoy@DMBR.UGent.be (F. van Roy).

0042-6822/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2007.01.044

The genes lacking the LTR had been designated as AL592309 and AL356957 (Illarionova et al., 2007), in correspondence to the GenBank accession numbers of the BAC clones encompassing the different genomic fragments. Based on our previous genomic analysis of the *NBPF* genes (Vandepoele et al., 2005), we annotated these genes as *NBPF3* and *NBPF17P*, respectively. Similarly, *NBPF* genes containing the LTR insertion had been designated AL049742, AL954711 and AL592284 (Illarionova et al., 2007), and we annotated them as *NBPF10*, *NBPF11* and *NBPF15*. RT–PCR experiments performed with different primer sets led Illarionova et al. (2007) to conclude that only *NBPF10*, *NBPF11* and *NBPF15* were expressed; the absence of RT–PCR products for *NBPF3* and *NBPF17P* led to the conclusion that these genes were not expressed.

We used the 5′ untranslated region (UTR) of the *NBPF11* transcript, as present in the KIAA1245 clone (nucleotides 1–770 from GenBank accession no. AB033071), in a BLAT search of the human genome (NCBI build 36.1). Many hits were obtained, corresponding to the regions previously described as harbouring *NBPF* genes (Vandepoele et al., 2005). The result for the *NBPF3* gene is shown in Fig. 1, which also shows the sequence of the *NBPF11* transcript, annotated as KIAA1245, and the regions that are homologous to the genomic locus of *NBPF3* (blocks connected by lines). The RepeatMasker track at the bottom of the figure clearly confirms that there is no LTR sequence in the genomic sequence flanked by the primers used by Illarionova et al. (2007). As expected, the genomic sequence surrounding the *NBPF3* transcription initiation site does contain repetitive elements, including LTRs, but we could not detect an HERV(K) LTR in this region.

In this analysis, we also observed numerous cDNA clones and ESTs annotated to this gene. These results corroborate our previous expression analysis of the *NBPF* genes (Vandepoele et al., 2005), in which we annotated many cDNAs to the *NBPF3* gene. The NBPF3 transcripts shown in Fig. 1 were derived from several tissues, both adult and embryonic, and cell lines including testis (17), brain (13) and embryonic stem cells (5). This demonstrates that *NBPF3* is expressed in different tissues.

Remarkably, none of the *NBPF3* transcripts contain the regions homologous to the second and third *NBPF11* exons, as present in the KIAA1245 sequence. We compared the genomic *NBPF3* sequences with the homologous exon–intron boundaries of the *NBPF11* gene, and found mutations in the *NBPF3* donor splice site for both ‘exons’ containing the primer-binding sites used by Illarionova et al. (2007). For exon 2, containing the primer-1 binding site, the TTT∗*GTAAGT* sequence in the *NBPF11* gene is mutated to TTT∗*GTAAAT* in *NBPF3*. For exon 3, containing the primer-3 and -4 binding sites, the AAG∗*GTGAGG* sequence in the *NBPF11* gene is mutated to AAG∗*TTGAGG* in *NBPF3* (asterisks denote real or putative exon–intron junctions; intronic sequence are in italics; mutated nucleotides are underlined). These mutations might interfere with the retention of these sequences as exons in the mature *NBPF3* transcript.

For the *NBPF17P* gene, although we identified a single DNA clone as a transcript from this gene (GenBank accession no. BE818279), we classified it as a pseudogene due to the presence of many nonsense mutations in the coding sequence (Vandepoele et al., 2005).

Apart from the intronic insertion of the HERV(K) LTR in the 5′ region, the *NBPF3* and *NBPF17P* genes differ from the other *NBPF* genes in the length of the intron between the coding exons of types 7 and 8. For these two genes, the intron is 3.2 kbp, whereas for other genes (*NBPF1*, *NBPF9-11*, *NBPF14-16* and *NBPF20*) it is approximately 3.9 kbp. As described earlier, the majority of the *NBPF* genes have a LINE repeat (Long INterspersed Element; L1P2) inserted in this intron (Vandepoele et al., 2005). The longer intron in the second group is due to the additional insertion of an L1PA4 repeat into the first LINE repeat. The presence or absence of the HERV(K) LTR and the L1PA4 LINE repeat clearly distinguishes the *NBPF3* and *NBPF17P* genes from the other *NBPF* genes and shows that these genes are at the evolutionary cradle of the *NBPF* gene family.

Discussion

The members of the *NBPF* gene family are located primarily on segmental duplications of chromosome 1 and have been subjected to several partial and complete gene duplications in recent evolution (Vandepoele et al., 2005). This evolution is still going on, as both copy number variation and structural variation between different individuals have been described for some *NBPF* genes (Redon et al., 2006; Tuzun et al., 2005).

It has been shown that after gene duplication, regulatory regions can diverge more rapidly than coding sequences, a phenomenon known as the duplication–degeneration–complementation model (Force et al., 1999). Apart from nucleotide mutations, the insertion of retroviral elements can also influence the transcriptional activities of the affected gene (Peaston et al., 2004; van de Lagemaat et al., 2003). During the evolution of the *NBPF* genes, an HERV(K) LTR insertion occurred in the 5′ region of one of the duplicated genes. Further duplications gave rise to the two recently described *NBPF* subfamilies (Illarionova et al., 2007). The limited number of nucleotide substitutions in different *NBPF* paralogs impedes a reliable phylogenetic analysis, but the analysis can be significantly improved by considering additional data. Here, we show that the *NBPF3* and *NBPF17P* genes, located on chromosomes 1p36 and 1q21, respectively, originated quite early during the evolution of the *NBPF* gene family, as they both lack the insertion of an

Fig. 1. Overview of the *NBPF3* locus. The 5′-UTR of KIAA1245 (nucleotides 1–770 of AB033071) was used in a BLAT search against the human genome, in combination with the forward (Pr1) and reverse (Pr3 and Pr4) primers used by Illarionova et al. (2007). The 5′-UTR of KIAA1245 shows homology to the genomic region of *NBPF3*, illustrated as blocks interconnected by lines, but none of the numerous transcripts from this gene contains the fragments to which the primers bind. The RepeatMasker track at the bottom clearly shows the absence of HERV(K) LTR in the genomic region flanked by Pr1 and Pr3. None of the repeats shown in the *NBPF3* locus are HERV(K) LTR sequences.