Available online at www.sciencedirect.com
SCIENCE DIRECT
Gene 366 (2006) 335–342
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

# Transcription of two human genes from a bidirectional endogenous retrovirus promoter

Catherine A. Dunn [a,b], Mark T. Romanish [a,b], Leanne E. Gutierrez [a,c], Louie N. van de Lagemaat [a,b], Dixie L. Mager [a,b,*]

[a] *Terry Fox Laboratory, BC Cancer Agency, 675 West 10th Avenue, Vancouver, BC, Canada V5Z 1L3*
[b] *Department of Medical Genetics, University of British Columbia, Vancouver, Canada*
[c] *Department of Biochemistry and Microbiology, University of Victoria, Canada*

Received 1 June 2005; received in revised form 29 August 2005; accepted 1 September 2005
Available online 8 November 2005
Received by C.W. Schmid

**Abstract**

Eight percent of the human genome is derived from endogenous retrovirus (ERV) insertions. ERV long terminal repeats (LTRs) contain strong promoters that are known to contribute to the transcriptional regulation of certain human genes. While some LTRs are known to possess bidirectional promoter activity in vitro, only sense orientation LTR promoters have previously been shown to regulate human gene expression.

Here we demonstrate that an ERV1 LTR acts as a bidirectional promoter for the human Down syndrome critical region 4 (*DSCR4*) and *DSCR8* genes. We show that while *DSCR4* and *DSCR8* are essentially co-expressed, their shared LTR promoter is more active in the sense than the antisense orientation. Through deletion analysis of the LTR we have identified positive and negative regulatory elements, and defined a core region of the promoter that is required for transcriptional activity in both orientations. Finally, we show that the ERV LTR also exists in the genomes of several non-human primates, and present evidence that potential transcription factor binding sites in the core region have been maintained throughout primate evolution.


*Keywords:* *DSCR4*; *DSCR8*; *MMA1*; *DCRB*; Long terminal repeat; Transposable element

## 1. Introduction

Down syndrome (DS) is the most common human chromosomal abnormality, present in approximately 1 in 800 live births. Most cases are caused by whole-chromosome trisomy 21, although partial and microtrisomies have also been reported (reviewed in (Antonarakis et al., 2004)). Comparison of the phenotypes and chromosome breakpoints of individuals with partial trisomies enabled identification of the DS critical region (DSCR). This 5.4 Mb region of chromosome 21q22.2, containing at least 33 genes (Gardiner et al., 2003), is thought to represent the minimal duplicated region required to reproduce some of the most common features of DS (Delabar et al., 1993). A recent mouse model study suggested that duplication of the DSCR alone may not be responsible for all aspects of the DS phenotype (Olson et al., 2004). However, it is likely that multiple genes within the DSCR contribute to the range of clinical traits observed in DS.

*DSCR4* and *DSCR8* are two genes of unknown function from the NCBI reference sequences collection (RefSeq, http://www.ncbi.nih.gov/RefSeq/). *DSCR4*, also known as *DCRB*, is expressed predominantly in the placenta (Nakamura et al., 1997). *DSCR8*, or malignant melanoma-associated 1 (*MMA1*), is thought to be a cancer-testis antigen, with expression restricted to the testis and certain malignant cells (de Wit et al.,

*Abbreviations:* *β3GAL-T5*, β1,3-galactosyltransferase 5; CRE, cAMP response element; DS, Down syndrome; DSCR, DS critical region; *EDNRB*, endothelin B receptor; ERV, endogenous retrovirus; HERV, human ERV; LTR, long terminal repeat; *MMA1*, malignant melanoma-associated 1; ORF, open reading frame; *PTN*, pleiotrophin; RACE, rapid amplification of cDNA ends; *SPAM1*, sperm adhesion molecule 1; TE, transposable element; TF, transcription factor.

* Corresponding author. Terry Fox Laboratory, BC Cancer Agency, 675 West 10th Avenue, Vancouver, BC, Canada V5Z 1L3. Tel.: +1 604 675 8139; fax: +1 604 877 0712.
*E-mail address:* dmager@bccrc.ca (D.L. Mager).

0378-1119/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2005.09.003

2002, 2005). Both *DSCR4* and *DSCR8* contain short predicted ORFs with no homology to any known protein (Nakamura et al., 1997; de Wit et al., 2005). The transcription start sites of *DSCR4* and *DSCR8* are located approximately 130 bp apart in a "head-to-head" orientation (Toyoda et al., 2002). This genomic arrangement is a common feature of gene pairs regulated by a bidirectional promoter (Adachi and Lieber, 2002; Takai and Jones, 2004; Trinklein et al., 2004). However, the putative shared *DSCR4*/*DSCR8* promoter region has not yet been functionally analyzed.

Transposable elements (TEs) comprise 45% of the human genome (Lander et al., 2001). Some TEs, notably endogenous retroviruses (ERVs), contain strong promoters that can contribute to the transcriptional regulation of human genes (reviewed in (Bannert and Kurth, 2004)). We performed a bioinformatic screen to identify human RefSeq genes that overlap TEs. We determined that *DSCR4* and *DSCR8* transcripts appear to initiate within the same ERV long terminal repeat (LTR), raising the possibility that the LTR acts as a bidirectional promoter for this human gene pair.

Some human ERV (HERV) LTRs are known to contain bidirectional promoter activity. For example, the LTRs of certain members of the HERV-E, -H, -K and -W families have been shown to activate transcription in both orientations in reporter gene assays (Feuchter and Mager, 1990; Domansky et al., 2000; Schulte et al., 2000; Lee et al., 2003). However, while many examples of human gene transcriptional regulation by HERV LTRs have been documented (van de Lagemaat et al., 2003; Bannert and Kurth, 2004), none involve reverse-orientation or bidirectional LTR promoter activity.

In this study, we show that transcription of *DSCR4* and *DSCR8* initiates within a shared ERV1 LTR promoter. RT-PCR analysis revealed that *DSCR4* and *DSCR8* are expressed in a similar range of normal human tissues. We determined that the ERV1 LTR contains strong bidirectional promoter activity, and identified the regions of the LTR responsible for transcription of *DSCR4* and *DSCR8*. Through comparative sequence analysis, we determined that a core region of the promoter required for bidirectional transcription has been maintained throughout primate evolution.

## 2. Materials and methods

### 2.1. Computational methods

Identification of human RefSeq transcripts that overlap TEs, sequence homology searches, examination of gene loci, comparison of cDNA and genomic DNA sequences, and identification of putative transcription factor (TF) binding sites were performed as described (Dunn et al., 2003; van de Lagemaat et al., 2003; Dunn and Mager, 2005). Macaque genomic DNA sequences were retrieved from the NCBI Sequencing Trace Archive (http://www.ncbi.nlm.nih.gov/Traces/trace.cgi). The sequence of macaque *DSCR8* splicing variant 3 has been submitted to GenBank with accession number DQ179114. Multiple sequence alignments were performed using the SeqWeb suite of programs (Accelrys).

### 2.2. Reverse transcription, PCR and Southern blotting

Normal human RNAs were purchased from Clontech; other RNAs were extracted from cultured cells using TRIzol (Invitrogen) according to the manufacturer's instructions. Five micrograms of each RNA was treated with DNaseI, then reverse transcribed and PCR amplified as described (Dunn et al., 2003). Primer sequences were as follows: *GAPDH*, as described (Dunn and Mager, 2005); *DSCR4,* forward (F) 5′-CTCACCCATGTGTGTCCATGTC (exon 1), reverse (R) 5′-CCCAACAGGATCCACAAGAAAC (exon 3); *DSCR8,* F 5′-CCGTCTTGTCCACTCGTTTCTG (exon 1), R 5′-GGAGTTCCAAGAAAAATTGTCACG (exon 4). Southern blotting was performed as described (Dunn et al., 2003). The *DSCR4* blot was probed with a $^{32}$P-labelled cDNA fragment containing sequence from exons 1, 2 and 3. The *DSCR8* blot was probed with a $^{32}$P-labelled oligonucleotide specific for exon 4 (5′-CATGCCAAGACCGAAAGTGCTG).

Non-human primate ERVs were amplified from genomic DNA as described (Dunn et al., in press), using the following primers: F 5′-GCTGGATTCCCTGGGAAGCTG (position 1-21 of the composite ERV element), R 5′-CAGCAAGATGAAGCTAAGGAAAC (non-repetitive sequence, 108 bp downstream of position 1148). The full sequences of the chimpanzee, gorilla, gibbon and baboon ERV-L/ERV1 elements have been submitted to GenBank with accession number DQ180329, DQ180330, DQ180331 DQ180332.

### 2.3. 5′-Rapid amplification of cDNA ends (RACE)

5′-RACE analysis of RNA from human placenta (*DSCR4*) or testis (*DSCR8*) was performed using the FirstChoice RLM-RACE kit (Ambion) as described (Dunn et al., 2003). Gene-specific primers were as follows: *DSCR4*, outer 5′-TCTCACATGGCCTCTCAGCAG, inner 5′-AACGTCAGCTTTGCTGCTGATC (both exon 3); *DSCR8*, outer 5′-GGAGTTCCAAGAAAAATTGTCACG, inner 5′-CATGAACAGCTTCATCATTCCAG (both exon 4).

### 2.4. Quantitative real-time RT-PCR

*DSCR8* transcripts were quantified using real-time RT-PCR, as described (Dunn et al., 2003). Dissociation curves were prepared to ensure that each primer pair amplified a single product. Serial dilutions of human testis RNA were used to prepare a standard curve for each primer pair. The level of *DSCR8* transcripts was normalized to the amount of *GAPDH* transcript in each tissue, and expressed relative to the low level of transcripts observed in the prostate. Primers were as follows: *GAPDH,* as described (Dunn and Mager, 2005); *DSCR8* F, 5′-CGTCTTGTCCACTCGTTTC (exon 1), R 5′-TTGGGTCCAGGCTCCTTC (exon 2).

### 2.5. Plasmid construction

The composite ERV-L/ERV1 element and the ERV1 LTR alone were amplified from human genomic DNA, as described