



## Intron 17 of the human retinoblastoma susceptibility gene encodes an actively transcribed G protein-coupled receptor gene.

H Herzog, K Darby, Y J Hort, et al.

*Genome Res.* 1996 6: 858-861

Access the most recent version at doi:[10.1101/gr.6.9.858](https://doi.org/10.1101/gr.6.9.858)

---

### References

This article cites 12 articles, 5 of which can be accessed free at:  
<http://genome.cshlp.org/content/6/9/858.full.html#ref-list-1>

### License

### Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

An advertisement banner with a teal background. On the left, the text reads "CRISPR and RNAi Genetic Screening. Your new superpower." In the center is a white box with the text "LEARN MORE". On the right is a woman wearing a red and white superhero cape and mask, with the Cellecta logo (a green molecular structure) and the word "CELLECTA" below it.

---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

Copyright © Cold Spring Harbor Laboratory Press

## RESEARCH

# Intron 17 of the Human Retinoblastoma Susceptibility Gene Encodes an Actively Transcribed G Protein-coupled Receptor Gene

Herbert Herzog,<sup>1</sup> Karen Darby, Yvonne J. Hort, and John Shine

Garvan Institute of Medical Research, St. Vincent's Hospital,  
Darlinghurst, Sydney, New South Wales 2010, Australia

The human retinoblastoma susceptibility gene, a member of the tumor suppressor gene family, is located on chromosome 13q14.12–13q14.2 and consists of 27 exons that are distributed over 180 kb. This study shows that intron 17, the largest in size, consisting of nearly 72,000 bp, contains an open reading frame encoding a novel G protein-coupled receptor in the reverse orientation relative to the transcription of the retinoblastoma susceptibility gene. Correction of a frameshift mutation revealed that this novel G protein-coupled receptor is the human homolog of a chicken T-cell-specific receptor cDNA. This is an additional description of an actively transcribed protein-encoding gene positioned within an intron of another gene, suggesting that introns can have important structural functions.

A common feature of eukaryotic genes is the presence of intervening sequences or introns, which must be excised from each RNA transcript to convert the precursor to an mRNA molecule that can code for the synthesis of a continuous protein. Introns range in size from <100 to >10,000 nucleotides and differ significantly from coding regions in exons, in that most of the nucleotide sequence of an intron can be altered without greatly affecting gene function. In addition, introns seem to have accumulated repetitive sequences and mutations rapidly during evolution, leading to the suggestion that they have no function at all and are largely genetic "junk." An exception to the view that introns do not encode functionally important sequences occurs in several genes specifying proteins involved in RNA processing and translation. Such genes often encode in their introns small nucleolar RNAs (snoRNAs), suggesting that cotranscription of snoRNAs with mRNAs for ribosomal or nucleolar proteins may provide a regulatory mechanism to coordinate accumulation of the different components required for ribosome biogenesis (Kiss et al. 1995; Tycowski et al. 1996).

Although the majority of eukaryotic genes contain introns, to date such sequences have not been found to add to the level of genetic complexity by containing other protein-encoding genes. Only a few genes are known that contain additional open reading frames (ORFs) in their introns, such as the neurofibromatosis type 1 gene (Cawthorn et al. 1990; O'Connell et al. 1990; Viskochil et al. 1991). We show here that an actively transcribed gene encoding a T-cell specific, G protein-coupled receptor is located within intron 17 of the human retinoblastoma (RB) susceptibility gene.

The human RB gene is located on chromosome 13q14.12–13q14.2 and consists of 27 exons that are distributed over 180 kb (Toguchida et al. 1993). The gene belongs to a family of tumor suppressor genes, and defects in this locus are responsible for retinoblastoma, a malignant tumor arising in the eyes of young children. Normally a 4.7-kb mRNA transcript is produced that encodes a 928 amino acid nuclear phosphoprotein. Of the 26 introns, intron 17 is the largest in size, consisting of nearly 72,000 bp.

In the process of using PCR with degenerative oligonucleotides to identify novel G protein-coupled receptor sequences we isolated a clone (U16) that proved to be the human homolog of a

<sup>1</sup>Corresponding author.  
E-MAIL [h.herzog@garvan.unsw.edu.au](mailto:h.herzog@garvan.unsw.edu.au); FAX 61 2 295 8 281.

## T-CELL SPECIFIC G PROTEIN-COUPLED RECEPTOR

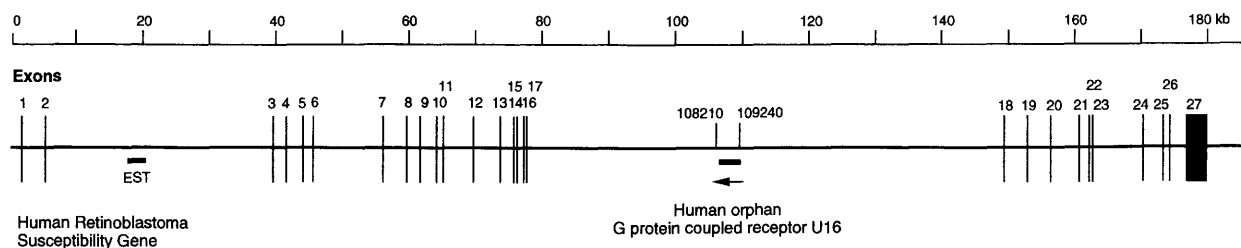
chicken T-cell-specific receptor cDNA (Kaplan et al. 1993). Further analysis of the human receptor sequence and searches of the GenBank/EMBL data bases revealed that an incomplete ORF encoding this gene exists within intron 17 of the RB gene. The direction of this ORF is in the reverse orientation relative to transcription of the RB gene. Translation of the upstream sequence of this putative gene in the two other reading frames revealed significant homology, including a translation initiation site and a conserved N-glycosylation site, to the chicken amino-terminal sequence, suggesting a possible sequencing error in the human RB gene sequence. This error was confirmed by a complete sequence analysis of PCR products covering the entire human G protein-coupled receptor-coding sequence (109,270–108,180), which were isolated from several individuals. All four fragments analyzed contained an additional thymidine nucleotide in position 109,145 corresponding to the RB gene sequence. Inclusion of this extra nucleotide joins the two reading frames. Furthermore, comparison of the human and chicken ORFs also indicated a possible sequencing error in the chicken sequence near the carboxy-terminal end. Removal of the adenine nucleotide in position 892 relative to the translation start site of the chicken 6H1 sequence increases significantly the identity between the two reading frames in this area. The overall amino acid identity between the human and the chicken receptor after correction of two frame-shift sequencing errors is 82%. The exact location of the human 344 amino acid ORF within the RB gene is between nucleotide 109,240 and 108,210 (Fig. 1). Comparison of human receptor sequence with other G protein-coupled receptor sequences shows the highest homology to purinergic, thrombin, and angiotensin receptors with 34%, 33.5%, and 30% identity, respectively (Vu et al. 1991; Curnow et al. 1992; Parr et al. 1994) (Fig. 2).

This is an additional description of an actively transcribed, protein-encoding gene positioned within an intron of another gene, suggesting that introns can have important structural functions and not only represent the remnant of past evolutionary events. There is no obvious functional connection between the RB gene and the accommodated T-cell-specific G protein-coupled receptor and no suggestion that these two genes are coregulated. The nested receptor ORF seems to be evolutionary well conserved as the human and chicken receptor sequences are highly homologous, suggesting an important role for this receptor.

The transcription of the human U16 mRNA, like the chicken 6R1 mRNA, seems to be limited to activated T cells (Kaplan et al. 1993). However, a number of T-cell stimuli, including concanavalin A and phytohemagglutinin, can strongly increase the level of the chicken 6R1 mRNA, which reaches its peak after 4 hr and is maintained over at  $\geq 45$  hr, suggesting an important role for this G protein-coupled receptor in activated T cells (Kaplan et al. 1993).

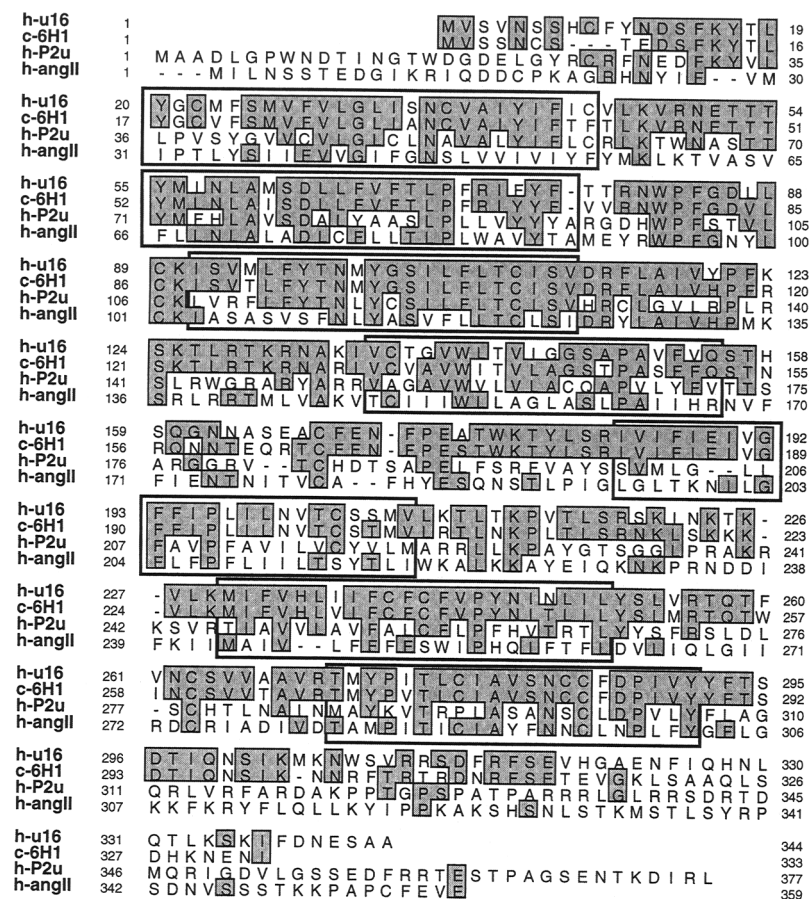
It is most likely that the human and chicken receptor genes arose from a gene duplication event via RNA-mediated transposition from an ancestral G protein-coupled receptor mRNA (Soares et al. 1985). Evolution of the G protein-coupled receptor gene superfamily probably involved several such events, as many family members lack introns, a somewhat unusual characteristic for mammalian genes (Probst et al. 1992). However, it might also be possible that the RB gene has been rearranged or extended during evolution and in that way incorporated the G protein-coupled receptor gene into intron 17.

The human RB gene contains a second large intron, intron 2, which is 33,894 bp in size. Two GpC islands in the middle of this intron also suggest the existence of possible additional ORFs. Sequence analysis and comparison with the Gen-



**Figure 1** Schematic representation of the human RB gene structure. The ORF in intron 17 encoding the human orphan G protein-coupled receptor U16 is indicated, together with the position of the EST fragment in intron 2.

## HERZOG ET AL.



**Figure 2** Comparison of the human G protein-coupled receptor U16 (h-u16), the human purinergic receptor (h-P2u), the human antiangiotensin II (h-angII), and the chicken T-cell-specific receptor (c-6R1) protein sequences. Highly conserved amino acids are indicated by shaded boxes. The putative *trans*-membrane domains are highlighted by bold frames.

Bank/EMBL data bank identified one 90% match with a 270-bp expressed sequence tag (EST) entry (accession no. F09341) derived from a human fetal brain cDNA library. Although at present there is no other evidence that this EST sequence represents a transcribed mRNA, the identification of the G protein-coupled receptor gene U16 within intron 17 of the human RB gene makes it very likely that other actively transcribed genes exist within the introns of other mammalian genes.

## ACKNOWLEDGMENTS

The nucleotide sequence reported in this paper has been submitted to the GenBank/EMBL data bank under accession no. L78805.

The publication costs of this article were defrayed in part by payment of page charges. This article must there-

fore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

## REFERENCES

- Cawthon, R.M., P. O'Connell, A.M. Buchberg, D. Viskochil, R.B. Weiss, M. Culver, J. Stevens, N.A. Jenkins, N.G. Copeland, and R. White. 1990. Identification and characterization of transcripts from the neurofibromatosis 1 region: The sequence and genomic structure of EV12 and mapping of other transcripts. *Genomics* **7**: 555-565.
- Curnow, K.M., L. Pascoe, and P.C. White. 1992. Genetic analysis of the human type-1 antidiuretic hormone receptor. *Mol. Endocrinol.* **6**: 1113-1118.
- Kaplan, M.H., D.I. Smith, and R.S. Sundick. 1993. Identification of a G protein coupled receptor induced in activated T cells. *J. Immunol.* **151**: 628-636.
- Kiss, T. and W. Filipowicz. 1995. Exonucleotide processing of small nucleolar RNAs from pre-mRNA introns. *Genes & Dev.* **9**: 1411-1424.
- O'Connell, P., D. Viskochil, A.M. Buchberg, J. Fountain, R.M. Cawthon, M. Culver, J. Stevens, D.C. Rich, D.H. Ledbetter, and M. Wallace. 1990. The human homolog of murine Evi-2 lies between two von Recklinghausen neurofibromatosis translocations. *Genomics* **7**: 547-554.
- Parr, C.E., D.M. Sullivan, A.M. Paradiso, E.R. Lazarowski, L.H. Burch, J.C. Olsen, L. Erb, G.A. Weisman, R.C. Boucher, and J.T. Turner. 1994. Cloning and expression of a human P2U nucleotide receptor, a target for cystic fibrosis pharmacotherapy. *Proc. Natl. Acad. Sci.* **91**: 13067-13071.
- Probst, W.C., L.A. Snyder, D.I. Schuster, J. Brosius, and S.C. Sealfon. 1992. Sequence alignment of the G-protein coupled receptor superfamily. *DNA Cell Biol.* **11**: 1-20.
- Soares, M.B., E. Schon, A. Henderson, S.K. Karathanasis, R. Cate, S. Zeitlin, J. Chirgwin, and A. Efstratiadis. 1985. RNA-mediated gene duplication: The rat preproinsulin I gene is a functional retroposon. *Mol. Cell. Biol.* **5**: 2090-2103.
- Toguchida, J., T.L. McGee, J.C. Paterson, J.R. Eagle, S. Tucker, D.W. Yandell, and T.P. Dryja. 1993. Complete genomic sequence of the human retinoblastoma susceptibility gene. *Genomics* **17**: 535-543.
- Tycowski, K.T., M.D. Shu, and J.A. Steitz. 1996. A

## T-CELL SPECIFIC G PROTEIN-COUPLED RECEPTOR

mammalian gene with introns instead of exons generating stable RNA products. *Nature* **379**: 464–466.

Viskochil, D., R. Cawthon, P. O'Connell, G.F. Xu, J. Stevens, M. Culver, J. Carey, and R. White. 1991. The gene encoding the oligodendrocyte-myelin glycoprotein is embedded within the neurofibromatosis type 1 gene. *Mol. Cell. Biol.* **11**: 906–912.

Vu, T.K., D.T. Hung, V.I. Wheaton, and S.R. Couglin. 1991. Molecular cloning of a functional thrombin receptor reveals a novel proteolytic mechanism of receptor activation. *Cell* **64**: 1057–1068.

*Received April 25, 1996; accepted in revised form June 21, 1996.*