Jpn J Human Genet 39, 393–401, 1994

GENOMIC STRUCTURE AND CHROMOSOMAL LOCALIZATION OF PROCESSED PSEUDOGENES FOR HUMAN RBP-J_k

Mingdi ZHANG,[1,2] Xiaoren TANG,[1] Chunlian JIN,[2]
Frederique LOGEAT,[3] Israël ALAIN,[3] Shinichi KONDO,[1]
Kailai SUN,[2] and Kazushige YOKOYAMA[1,*]

[1]*Tsukuba Life Science Center, The Institute of Physical and Chemical Research (RIKEN), 3-1-1 Koyadai, Tsukuba, Ibaraki 305, Japan*
[2]*China Medical University, Shenyang P. R. China*
[3]*Unité Biologie Moléculaire de l'Expression Génique, Department Biologie Moléculaire, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cédex 15, France*

Summary The functional gene for human recombination signal sequence-binidng protein (RBP-J_k) and corresponding processed psudogenes have been isolated from various species, such as *Drosophila, Xenopus,* mouse, and human. Here we report the isolation of another two genomic pseudogenes of human RBP-J_k, named K2 and K7, from a cosmid library of Hela cells. The nucleotide sequences of both genes exhibited more than 95% homology to the functional human gene for RBP-J_k. Moreover, they did not contain any intron sequences and were interrupted by several stop codons in all frames. *In situ* hybridization demonstrated that the pseudogenes, K2 and K7, were localized at chromosomes 9p13 and 9q13, respectively. Their physical maps differed from those of the true functional gene and of the pseudogenes reported previously by Amakawa *et al.* (1993).

Key Words RBP-J_k, pseudogene, cosmid, *in situ* hybridization, physcal map

INTRODUCTION

The variable (V), diversity (D), and joining (J) recombinations of immunoglobulin genes require conserved recombination signal sequences (RS), namely, a heptamer CACTGTG and a T-rich nonamer GGTTTTTGT, which are separated by a spacer sequence of either 12 or 23 nucleotides (Tonegawa, 1983; Honjo and

Received July 18, 1994; Revised version accepted August 15, 1994.
* To whom all correspondence should be addressed.

Habu, 1985; Korenberg *et al.*, 1986). The enzymes involved in V-(D)-J recombination as recombinases should recognize the conserved signal sequences, catalyzing site-specific cleavage and ligation of cleaved ends. To date, four genes have been suggested to be involved in the V-(D)-J recombination. These are RAG-1 (Mombaerts *et al.*, 1992), RAG-2 (Shinkai *et al.*, 1992), the *scid* gene (Lieber *et al.*, 1988; Fulop and Philips, 1990) and RBP-J_k. The RBP-J_k gene has been identified in various species, such as *Drosophila*, mouse, and human (Furukawa *et al.*, 1991, 1992; Matsunami *et al.*, 1989; Kawaichi *et al.*, 1992; Amakawa *et al.*, 1993), as have analogous pseudogenes (Amakawa *et al.*, 1993).

The RBP-J_k gene has been shown to have been highly conserved during evolution from *Drosophila* to human (Furukawa *et al.*, 1991; Amakawa *et al.* 1993). The *Drosophila* RBP-J_k protein is 75% identical to its murine counterpart (Furukawa *et al.*, 1991). The human and mouse RBP-J_k proteins are 98% identical (Amakawa *et al.*, 1993). These proteins have molecular masses of 60 kDa and contain a region of 40 amino acids that is homologous to the integrase motif (Hamaguchi *et al.*, 1989; Matsunami *et al.*, 1989). Recent studies demonstrated the specific binding of mouse RBP-J_k and human RBP-2N to the sequences 5′-CGTGGGAA-3′ and 5′-TGGGAAAGAA-3′, respectively (Tun *et al.*, 1994; Dou *et al.*, 1994). The function of this protein in recombination remains to be ascertained. However, the *Drosophila* homolog of RBP-J_k was found to be the *Suppressor* of *Hairless* (Furukawa *et al.*, 1991, 1992; Schweisguth and Posakony, 1992), which participates in the regulation of neurogenic expression, an observation that suggests that RBP-J_k may play a role in the regulation of gene expression. Dou *et al.* (1994) recently reported that RBP-2N may have a general role in transcriptional repression. Two types of processed pseudogene for human RBP-J_k have been isolated and an evolutionary mechanism for their generation has been proposed (Amakawa *et al.*, 1993). However we don't know how many genes are present in human chromosomes. Moreover the genomic organization of these genes included pseudogenes are unknown. In order to understand the structural organization of RBP-J_k, we have isolated and characterized the genomic clones of RBP-J_k. We report here the genomic structure and the chromosomal location of two newly identified pseudogenes, K2 and K7, from a cosmid library of Hela cells. These pseudogenes are different from the pseudogenes reported by Amakawa *et al.* (1993).

MATERIALS AND METHODS

Screening of a human genomic DNA library. A cosmid library was constructed from the genomic DNA of Hela cells, which was partially digested with *Sau*3A and then ligated to the cosmid vector pWE15 (Toyobo, Tokyo). The screening of the library (a total of 2.5×10^6 cosmid clones was screened) by colony hybridization with a probe of 1.7-kb *Eco*RI digested DNA fragment of RBP-J_k

(Sambrook *et al.*, 1989; Amakawa *et al.*, 1993; Feinberg and Vogelstein, 1983).

Nucleotide sequencing. The cosmid cloned DNAs were sequenced directly (Slightom and Sieu, 1992) by a modified method as described by Tang *et al.* (1993).

Restriction maps of cosmid clones. The restriction maps of the cosmid-cloned K2 and K7 DNA were generated by a new method. DNAs were completely digested with *Eco*RI and their restriction fragments were recovered. Each recovered fragment of DNA was subcloned into the *Eco*RI-digested pBluescript SK^+ vector (Stratagene, La Jolla, CA). Each subcloned DNA was purified and sequenced from both flanking regions of each DNA insert with M13 universal primers. Oligonucleotide DNAs as reverse primers were digested on the basis of the nucleotide sequence of each subclone from the 3′ to the 5′ direction and were synthesized on a DNA synthesizer (model 380B; Applied Biosystems, Foster City, CA). Then, the cosmid DNA of K2 or K7 was directly sequenced using the primers. A search was made for homology between the nucleotide sequences of cosmid clone K2 or K7 generated with reverse primers and those of the respective subclones generated with M13 universal primers to identify linked subclones. If the nucleotide sequence generated by a reverse primer from a given cosmid clone was homologous to the nucleotide sequence generated by a M13 universal primer from one of the subclones, this subclone should be linked to another subclone containing the nucleotide sequence of the reverse primer that was used for directly sequencing of the previous cosmid clone. Likewise, the linked subclones containing different *Eco*RI DNA fragments of K2 or K7 were ordered.

In situ chromosome hybridization. The preparation of biotin labeled probe of cosmid DNAs, *in situ* chromosome hybridization and G-banding were carried out as described previously (Lichter *et al.*, 1988, 1990; Zabel *et al.*, 1983).

RESULTS AND DISCUSSION

In the course of screening a Hela cell cosmid library for human RBP-J_k genes, we obtained two clones, K2 and K7, which hybridized specifically with human RBP-J_k cDNA clone PCR-3 (Amakawa *et al.*, 1993). As shown in Fig. 1a, the *Eco*RI-digested DNA fragments of K2 were different from those of K7. They were also different from those of the functional human RBP-J_k gene and the pseudogenes reported by Amakawa *et al.* (1993). The DNA fragments of K2 and K7 that had been digested with *Eco*RI were allowed to hybridize with probe c, derived from mouse RBP-J_k DNA, as described by Amakawa *et al.* (1993). The results (see Fig. 1b) showed that the sizes of hybridized DNA fragments of K2 (3.8 and 1.4 kb) and K7 (3.8 and 1.7 kb) were different from those of the human RBP-J_k gene (18.0, 3.8, 3.5, 2.5, and 2.1 kb) and those of two known pseudogenes (one, 3.5 and 2.7 kb; the other, 3.8 and 2.7 kb; Amakawa *et al.* (1993). The *Eco*RI restriction maps of K2 and K7 were compared with the physical maps of the human RBP-J_k gene and the corresponding pseudogenes (Amakawa *et al.*, 1993) and it

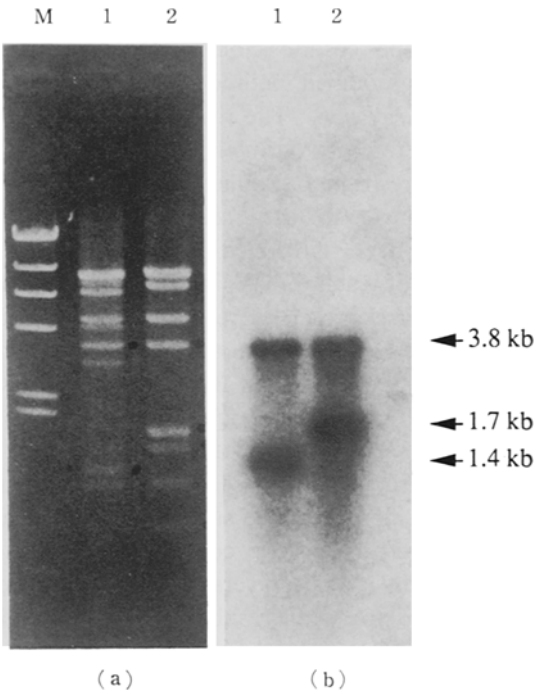

Fig. 1. (a) Results of agarose gel (1%) electrophoresis of *Eco*RI-digested genomic DNA from K2 (lane 1) and K7 (lane 2). Lanes marked M contain fragments of lambda DNA that had been digested with *Hin*dIII. (b) Southern blot analysis of *Eco*RI-digested genomic DNA from K2 and K7 with probe c from the mouse gene for RBP-J$_\kappa$ (Amakawa *et al.*, 1993). The hybridized DNA fragments are indicated by arrowheads (1.4, 3.8 kb, lane 1; 1.7, 3.8 kb, lane 2).

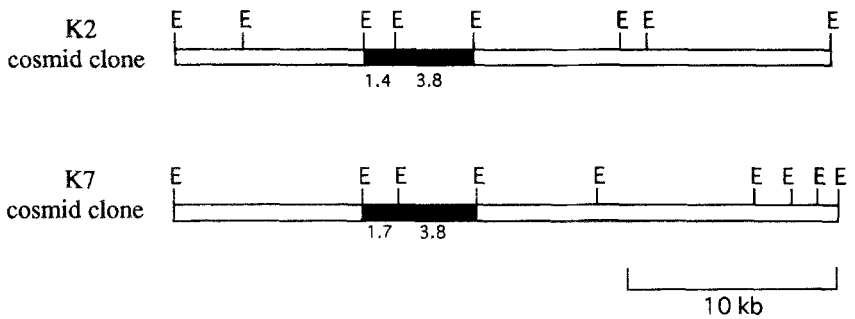

Fig. 2. Genomic structures of pseudogenes K2 and K7. E indicates a restriction site of *Eco*RI. Solid boxes indicate the regions specific to probe c from the mouse gene for RBP-J$_\kappa$ (Amakawa *et al.*, 1993).

was concluded that they were different clones (see Fig. 2). These data indicated that cosmid clones K2 and K7 were new genomic clones of human RBP-J$_k$, isolated from a cosmid library of Hela cells.

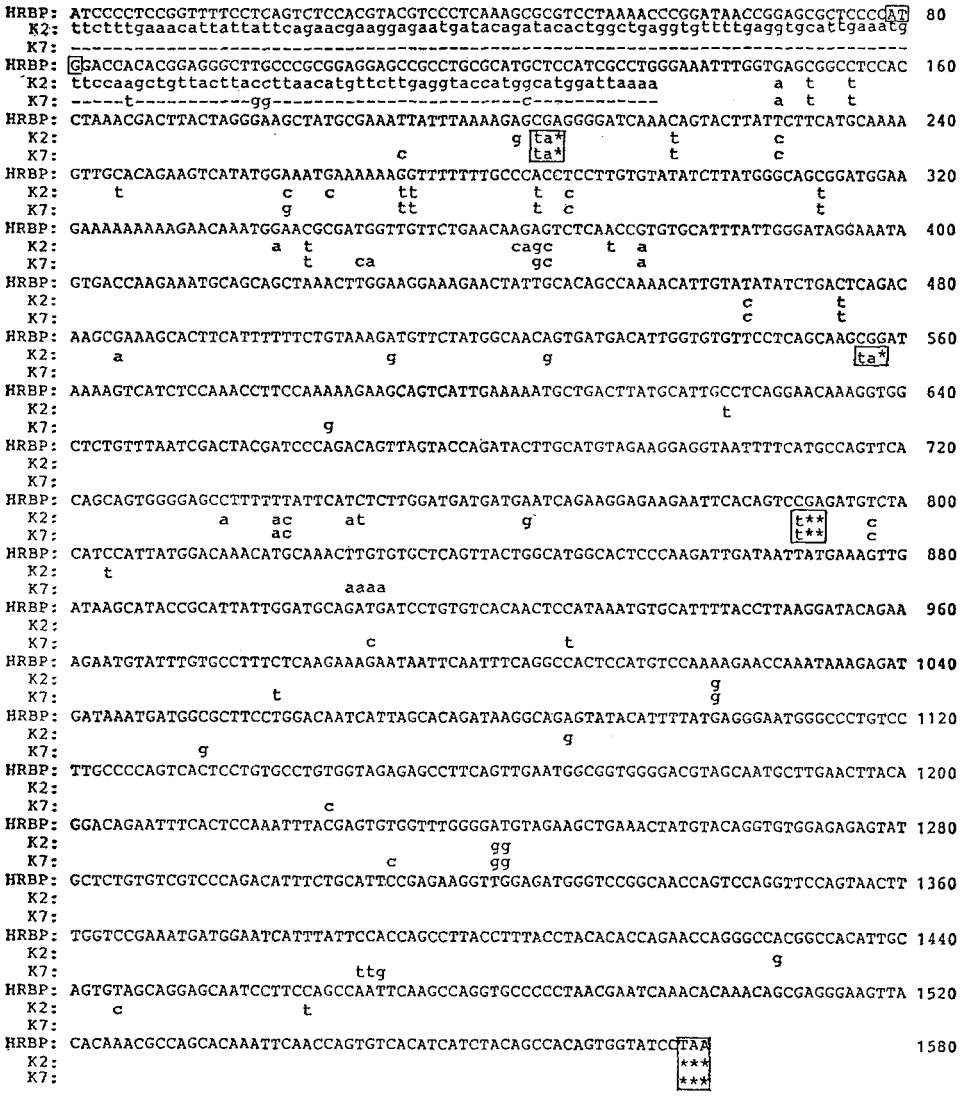

Fig. 3. Comparison of the sequence of the human cDNA for RBP-J$_k$ (HRBP; Amakawa *et al.*, 1993) with those of the processed pseudogenes (K2 and K7) isolated in this study. Only non-homologous nucleotides in the sequences of K2 or K7 are shown. Mutated stop codons of K2 or K7 are indicated by an open box with asterisks. The translocation start site and the stop codon of HRBP are indicated by an open box. Nucleotide sequences of K2 and K7 have been deposited in GenBank under accession numbers L34543 and L34544, respectively.

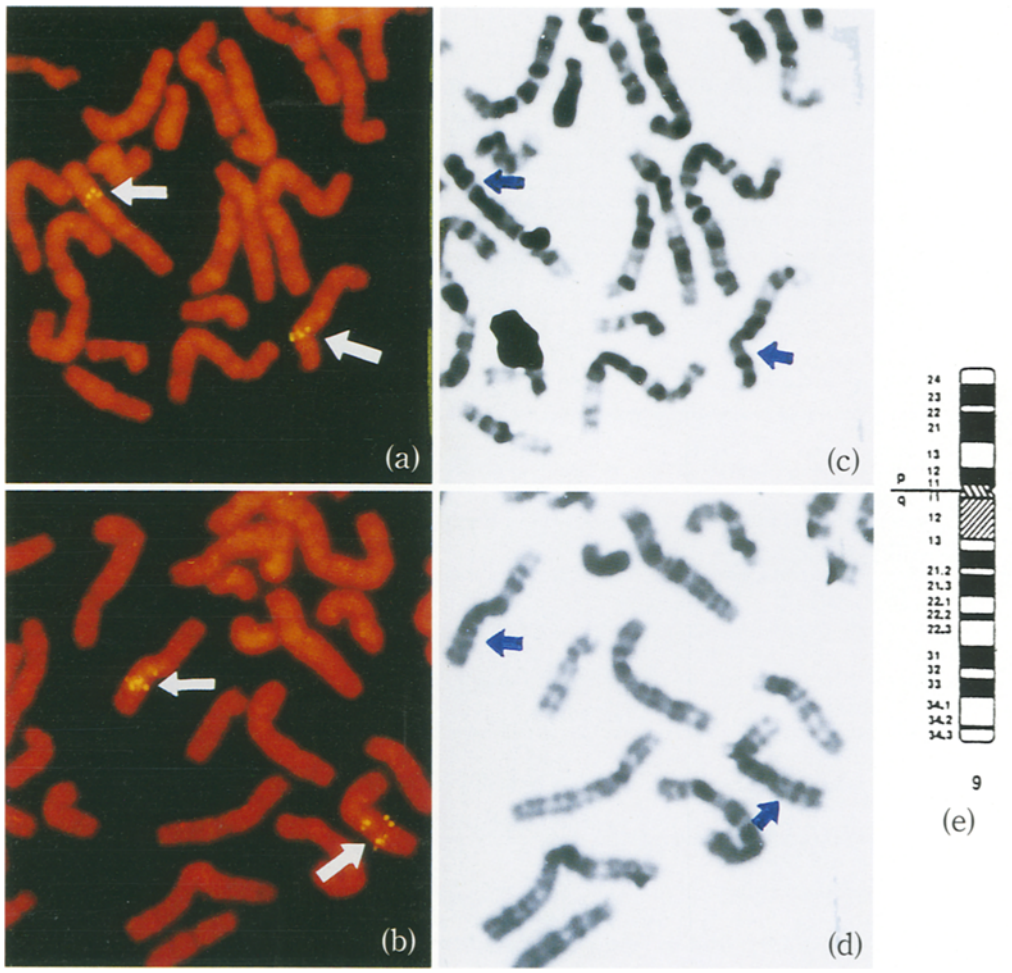

Fig. 4. *In situ* hybridization on human chromosomes with a biotinylated fragment of K2 (a) or K7 (b) genomic DNA as probe. (a and b) Indirect immunofluorescence detection with propidium iodide counterstaining. (c and d) G-banding of the same chromosome as those shown in (a) and (b). The fluorescent and G-banding spots, indicated by arrows, are located at chromosome 9p13 and 9q13 (e).

In an attempt to determine the nucleotide sequences of both the K2 and the K7 clone, we synthesized several primers that were based on the known nucleotide sequence of the human RBP-J_k gene (Amakawa *et al.*, 1993). First, five oligonucleotide reverse primers were designed, on the basis of the complementary sequence of the human RBP-J_k gene from the 3′ to 5′ flanking region, and synthesized, as follows: 1, 5′-GCT GTA GAT GAT GT (positions 1,565–1,552 of the sequence of the human RBP-J_k gene, accession number: L07872); 2, 5′-TTC TTG

AGA AAG GC (positions 987–974); 3, 5′-AAG GTT TGG AGA TG (positions 580–567); 4, 5′-CTG CCC ATA AGA TA (positions 311–298); and 5, 5′-ACC AGT GTC ACA TC (positions 1,543–1,556). The cosmid DNA of K2 or K7 was then sequenced directly using these primers. As shown in Fig. 3, both K2 and K7 exhibited 95% homology to the human RBP-J_k gene. A total of over forty mutations, including five (K2) and four (K7) stop codons, were observed and neither sequence contained any introns. We also isolated cDNA clones that corresponded to K2 and K7 by using the oligonucleotide probes of coding sequence of HRBP cDNA (data not shown). Thus, we suggest that both cosmid clones K2 and K7 represent new processed pseudogenes of human RBP-J_k.

We next examined the chromosomal location of K2 and K7 genes by *in situ* hybridization (Fig. 4). Intense signals from both K2 and K7 were detected at chromosome 9p13 and 9q13 (Figs. 4, a, b, c, d, and e). As reported by Amakawa *et al.* (1993), the functional human gene for RBP-J_k was assigned to chromosome 3q25 and two pseudogenes 1 and 2 were mapped to 9p13 and 9q13 respectively. They suggested that multiple copies of pseudogene 1 may exist at 9p13, and the single copy of pseudogene 2 may exist on 9q13. However we detected multiple copies of the RBP-J_k gene in two regions although the signal of 9p13 was more intense than that of 9q13. The locations of the functional gene and the pseudogenes may differ for the following reasons. (1) The large introns interspersed within the functional gene as reported by Amakawa *et al.* (1993), of which DNA sequences different from those of the pseudogenes, could contribute to signals at different loci after *in situ* hybridization. (2) The human functional gene for RBP-J_k may have been evolutionarily conserved at chromosome 3q25 during duplication and the pseudogene, without introns, may have been generated by recombination mutation and/or translocation to chromosome 9p13 and 9q13. Chromosomal inversion between 9p13 and 9q13, with subsequent divergence by point mutations, nucleotide insertions and deletions, may have occurred during this duplication. More than four or at least four different genomic pseudogenes of human RBP-J_k may be present on human chromosomes. These pseudogenes did not contain any intron sequences. We should now try to isolate the remaining pseudogenes for human RBP-J_k, if they exist, and characterize their structures for a comparison with the true gene so that we can attempt to explain the evolutionary mechanism of generation of pseudogenes.

Acknowledgments The authors thank Dr. T. Honjo for the cDNA clone for mouse RBP-J_k. This work was supported by Special Coordination Funds of the Science and Technology Agency of the Japanese Government.

REFERENCES

Amakawa R, Wu J, Ozawa K, Matsunami N, Hamaguchi Y, Matsuda F, Kawaichi M, Honjo T (1993): Human J_k recombination signal binding protein gene (IGKJRB): comparison with

its mouse homologue. Genomics **17**: 306–315

Dou S, Zeng X, Patricia C, Hediye E-B, Paul T, Honjo T, Lynne DV (1994): The recombination signal sequence-binding protein RBP-2N functions as a transcriptional repressor. Mol Cell Biol **14**: 3310–3319

Feinberg AP, Vogelstein B (1983): A technique for radiolabeling DNA restriction fragments to high specific activity. Anal Biochem **132**: 6–13

Fulop GM, Philips RA (1990): The acid mutation in mice causes a general defect in DNA repair. Nature **347**: 479–482

Furukawa T, Kawaichi M, Matsunami N, Ryo H, Hishida Y, Honjo T (1991): The *Drosophila* RBP-J_k gene encodes the binding portein for the immunoglobulin J_k recombination signal sequence. J Biol Chem **266**: 23334–23340

Furukawa T, Maruyama M, Kawaichi M, Honjo T (1992): The *Drosophila* homolog of the immunoglobulin recombination signal-binding protein regulates peripheral nervous system development. Cell **69**: 1191–1197

Hamaguchi Y, Matsunami N, Yamamoto Y, Honjo T (1989): Purification and characterization of a protein that binds recombination signal sequence of the immunoglobulin J_k segment. Nucleic Acids Res **17**: 9015–9026

Honjo T, Habu S (1985): Origin of immune diversity: genetic variation and selection. Annu Rev Biochem **54**: 803–830

Kawaichi M, Oka C, Shibayama S, Koromilas AE, Matsunami N, Hamaguchi Y, Honjo T(1992): Genomic organization of mouse J_k recombination signal-binding protein (RBP-J_k) uəəe. J Biol Chem **267**: 4016–4022

Korenberg M, Sui G, Hood L, Shastri N (1986): The molecular genetics of the T-cell antigen receptor and T-cell antigen recognition. Annu Rev Immunol **4**: 529–591

Lichter P, Cremer T, Tang C-JC, Watkins PC, Manuelidis L, Ward DC (1988): Rapid detection of human chromosome 21 aberrations by *in situ* hybridization. Proc Natl Acad Sci USA **85**: 9664–9668

Lichter P, Chang-Tang CJ, Call K, Hermanson G, Evans GA, Housman D, Ward DC (1990): High-resolution mapping of human chromosome 11 by *in situ* hybridization with cosmid clones. Science **247**: 64–69

Lieber MR, Hesse JE, Lewis S, Bosma GC, Rosenberg N, Mizuuchi K, Bosma MG, Gellert M (1988): The defect in mouse severe combined immune deficiency: joining of signal sequences but not coding segments in V(D)J recombination. Cell **55**: 7–16

Matsunami N, Hamaguchi Y, Yamamoto Y, Kuze K, Kangawa K, Matsuo H, Kawaichi M, Honjo T (1989): A protein binding to the J_k recombination sequence of immunoglobulin genes contains a sequence related to the integrase motif. Nature **234**: 934–937

Mombaerts P, Iacomini J, Johnson RS, Herrup K, Tonegawa S, Papaioannuou VE (1992): RAG-1 deficient mice have no mature B and T lymphocytes. Cell **68**: 869–877

Sambrook J, Fritsch EF, Maniatis T (1989): Molecular cloning: A laboratory manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Schweisguth F, Posakony JW (1992): *Suppressor* of *Hairless*, the *Drosophila* homolog of the mouse recombination signal-binding protein gene, controls sensory organ cell fates. Cell **69**: 1199–1212

Shinkai Y, Rathbun G, Lam KP, Oltz EM, Stewart V, Mendelsohn M, Charron J, Datta M, Young F, Stall AM, Alt FW (1992): RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. Cell **68**: 855–867

Slightom JL, Sieu LC (1992): Direct sequencing of baculovirus genomic DNA: sequence determination of the engineered respiratory syncytial virus chimeric FG gene. BioTedhniques **13**: 94–105

Tang X, Wang Y, Nakata Y, Li H, Fujita A, Gao H, Sarai A, Yokoyama K (1993): A mathema-

tically designed STS primer without any mismatches for direct sequencing of cosmid DNA clones. Jpn J Human Genet **38**: 381–390

Tonegawa S (1983): Somatic generation of antibody diversity. Nature **302**: 575–581

Tun T, Hamaguchi Y, Matsunami N, Furukawa T, Honjo T, Kawaichi M (1944): Recognition sequence of a highly conserved DNA binding protein RBP-J_K. Nucleic Acids Res **22**: 965–971

Zabel BU, Naylor SL, Sakaguchi AY, Bell GI, Shows TB (1983): High resolution of human genes for amylase, proopiomelanocortin, somatostatin, and a DNA fragment (D3S1) by *in situ* hybridization. Proc Natl Acad Sci USA **80**: 6932–6936