Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 9131–9134, December 1987
Immunology

Close linkage of the mouse and human CD3 γ- and δ-chain genes suggests that their transcription is controlled by common regulatory elements

(T-cell antigen receptor/human chromosome 11/mouse chromosome 9/gene duplication)

HARUO SAITO*, TAKUMI KOYAMA*, KATIA GEORGOPOULOS†, HANS CLEVERS†, WAYNE G. HASER‡, TUCKER LEBIEN§, SUSUMU TONEGAWA‡, AND COX TERHORST†

*Division of Tumor Immunology and †Laboratory of Molecular Immunology, Dana–Farber Cancer Institute, Harvard Medical School, Boston, MA 02115; ‡Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139; and §Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN 55455

Contributed by Susumu Tonegawa, July 27, 1987

ABSTRACT **Antigen receptors on the T-cell surface are noncovalently associated with at least four invariant polypeptide chains, CD3-γ, -δ, -ε, and -ζ. The mouse CD3-γ gene, consisting of seven exons, was found to be highly homologous to the CD3-δ gene described earlier. Both the high level of sequence homology and the exon/intron organization indicate that the CD3-γ and -δ genes arose by gene duplication. Surprisingly, murine and human genomic DNA clones could be isolated that contained elements of both the CD3-γ and CD3-δ genes. In fact, the putative transcription start site of the mouse CD3-γ gene is less than 1.4 kilobases from the transcription initiation site of the mouse CD3-δ gene. Common elements that regulate the divergent transcription of the two genes are therefore proposed to be located in the intervening 1.4-kilobase DNA segment. This might contribute to the coordinate expression of the CD3-γ and -δ genes during intrathymic maturation of T lymphocytes.**

Several types of T-cell receptors (TCRs) for antigen (and major histocompatibility complex proteins) have been described (1, 2). In addition to variable proteins (TCR-α, -β, -γ, or -δ), antigen receptors contain a set of invariant proteins—CD3-γ, -δ, -ε, and -ζ (3–5)—when isolated from the plasma membrane of T lymphocytes. A fifth, intracellular protein, CD3-ω, binds transiently with members of the complex during biosynthesis (6, 7). Expression of the TCR and CD3 genes appears to be confined to thymus-derived lymphocytes. CD3 mRNA and cytoplasmic protein are detected in the earliest recognizable thymocytes, whereas transcription of TCR-α, -β, -γ occurs after further maturation within the thymus gland (8–10). A detailed study of the CD3 genes and the regulation of their tissues-specific expression has, therefore, been undertaken. Earlier, the genes coding for the CD3-δ and -ε proteins were isolated and mapped to band q23 on the human chromosome 11 and to mouse chromosome 9 (refs. 11, 13, and 14; H.C., S. Dunlap, H.S., T. Wileman, and C.T., unpublished results). Here we report the cloning of the mouse CD3-γ gene¶ and demonstrate that the CD3-γ and CD3-δ genes are closely linked. The putative 5′ end of the mouse CD3-γ gene was located 1.4 kilobases (kb) from the transcription start site of the CD3-δ gene. Transcription of the two genes therefore takes place in opposite directions. Since a close linkage was also found for the human CD3-γ and CD3-δ genes, the existence of common transcription regulation factors acting on sequences present in the DNA segment between the two genes is highly likely.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

For isolation, subcloning, and sequence analysis of the genomic DNA, published methods were used (11, 15). The mouse cDNA clone pT134 (16) was used as a probe to screen a BALB/c liver genomic DNA library made with partial *Mbo* I fragments in bacteriophage EMBL3. Positive bacteriophages λT3-γ1 (insert size 11.5 kb) and λT3-γ2 (insert size 12 kb) were isolated, the exon-containing fragments were subcloned in plasmid vectors, and nucleotide sequences were determined according to Maxam and Gilbert (17).

A 0.7-kb *Eco*RI–*Acc* I fragment from pTM-δ2 (originally called pMT-2, ref. 11) was subcloned in bacteriophage M13 mp8 or mp9. Single-stranded templates were prepared from these subclones and sequence determination was done by the dideoxy chain-termination procedure of Sanger *et al.* (18).

From a genomic DNA library made with a partial *Mbo* I digest of human leukocyte DNA and the bacteriophage EMBL3, 5×10^5 bacteriophages were screened with the human CD3-γ cDNA pJ6T3-γ-2 (19) as a probe after labeling with [α-^{32}P]dGTP by the random-priming technique (20). Of the two bacteriophages thus isolated, one, λTWL-γ1, hybridized with a 26-mer synthetic oligodeoxynucleotide complementary to the 5′ end of that CD3-γ cDNA (GCCAGGCCCTTCCCCTGTTCCATGTC) (19) and therefore contained the first exon of the human CD3-γ gene. λTWL-γ1 was then used for restriction enzyme mapping by standard techniques (12). In addition to the two probes mentioned above, the human CD3-δ cDNA pPBGC-9 (11) and a human CD3-δ genomic DNA segment covering the transcription start site [made by cleavage with *Hin*dIII and *Bst*EII (at positions −354 and +57, respectively)] were used.

RESULTS AND DISCUSSION

Among a collection of T-lymphocyte-specific cDNA clones isolated by subtractive hybridization techniques (21), group K (22), which included nine independently isolated clones, was identified as the mouse counterpart (16) of the human CD3-γ cDNA after comparison of the respective nucleotide sequences (19). As was shown with the human CD3-γ and CD3-δ polypeptide chains, the primary amino acid sequence of the mouse CD3-γ chain (16) deduced from the nucleotide sequence showed strong homologies with the mouse CD3-δ chain (23). To further characterize the evolutionary relationship of these two genes, we isolated genomic DNA clones

Abbreviation: TCR, T-cell receptor.

¶The sequence of the mouse CD3-γ gene is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03590).

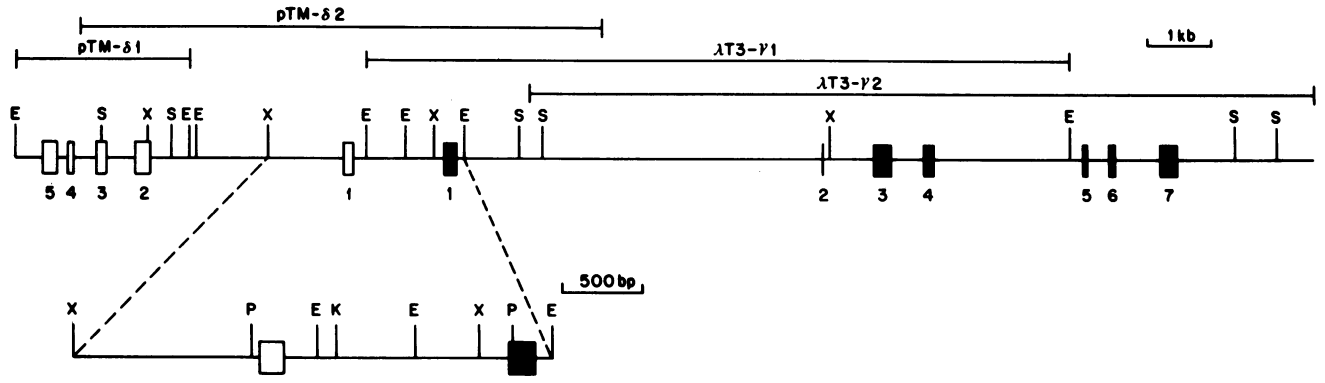

FIG. 1. Partial restriction map of the area covering the mouse CD3-γ and CD3-δ genes. The mouse CD3-γ gene spans 14.5 kb and contains seven exons (filled boxes). The mouse CD3-δ gene was found to span 5 kb and contain five exons (open boxes) (11). Bacteriophages λT3-γ1 and λT3-γ2, which contain the mouse CD3-γ gene, were isolated from a genomic DNA library made from BALB/c liver DNA. pTM-δ1 and pTM-δ2 are plasmids containing a 2.7-kb *Eco*RI fragment and an 8-kb *Pst* I fragment of genomic DNA, respectively (11). Restriction enzyme sites are indicated as follows: E, *Eco*RI; K, *Kpn* I; P, *Pvu* II; X, *Xba* I; S, *Stu* I. bp, Base pairs.

containing the mouse CD3-γ gene from a BALB/c mouse liver DNA library. Two overlapping bacteriophage clones, λT3-γ-1 and λT3-γ-2, spanned a 14.5-kb segment of mouse genomic DNA and contained the entire CD3-γ gene (Fig. 1). By a combination of restriction enzyme mapping, Southern blot (24) hybridization, and comparison of the nucleotide sequence with the CD3-γ cDNA sequence (16), seven exons encoding the mouse CD3-γ chain were detected (indicated schematically in Fig. 1).

Several conclusions can be drawn from the nucleotide sequences of the seven CD3-γ exons and their flanking regions (Fig. 2). A comparison of the exons encoding the CD3-γ chain and the five exons encoding the mouse CD3-δ chain (11) indicated that the exon/intron organization (i.e., the size of each exon and the codon usage in the splice sites) of the two genes was very similar (Table 1). The major differences are that the CD3-γ gene contains a 24-bp exon (exon 2), which does not have a counterpart in the CD3-δ gene, and that exon 3 of the CD3-γ gene contains a 9-bp insertion as compared to exon 2 of CD3-δ gene. Moreover, the mouse CD3-γ gene contains a separate exon encoding the 3′ untranslated region. The similarity in exon/intron organization as well as the high level of primary sequence homology (19) between the CD3-γ and -δ genes strongly suggests that these two genes arose by a gene-duplication event.

Because the CD3-γ and -δ genes appear to be expressed coordinately in all thymus-derived lymphocytes (8–10), we searched for similarities of putative regulatory sequences in 5′ regions of the CD3-γ and -δ genes. Surprisingly, we found that about 350 bp of the 5′ sequence of the mouse CD3-γ gene were perfectly complementary to the 5′ sequence of the mouse CD3-δ gene. By comparing the restriction maps of the CD3-γ and -δ genes, the 5′-most *Eco*RI site of the λT3-γ-1 clone was found to coincide with the *Eco*RI site up to the 5′ of the first exon of the mouse CD3-δ gene (Fig. 1). Although the precise transcription initiation site of the CD3-γ gene has not been determined, a comparison of the genomic and cDNA sequences indicated that the initiation site is no more than 1175 bp from the *Eco*RI site. Since the transcription initiation site of the CD3-δ gene is within 200 bp (11) on the other side of the *Eco*RI site, the distance between the two genes is not more than 1400 bp. This finding suggests (*i*) that regulation of expression of the mouse CD3-γ and CD3-δ genes may indeed be coordinated via elements located in the intervening segment between their respective promoters and (*ii*) that transcription occurs in opposite directions.

Having determined the close proximity of the mouse CD3-γ and -δ genes, we then investigated a possible close linkage between the human genes. A human CD3-γ cDNA clone (19) was used to isolate bacteriophages coding for the CD3-γ gene from a genomic DNA library made from human lymphocytes. To ensure the presence of the 5′-most exon of CD3-γ gene in the bacteriophage isolates, the positive genomic DNA clones were further analyzed with a synthetic oligonucleotide probe complementary to the 5′ end of the human CD3-γ cDNA (GCCAGGCCCTTCCCCTGTTCCATGTC) (19). One bacteriophage, λTWL-γ1, was thus selected and its ≈20-kb-long insert was mapped by restriction enzyme analysis and Southern blotting with the human CD3-γ cDNA pJ6T3γ-2 (19), the 5′ synthetic oligonucleotide, the human CD3-δ cDNA pPBGC9 (25), and a genomic DNA segment covering the human CD3-δ transcription initiation site (11) used as hybridization probes. As shown in Fig. 3, the first exon of the human CD3-δ gene and the putative first exon of CD3-γ were located on a 2.6-kb *Sal* I–*Bam*HI fragment of the 20-kb insert in λTWL-γ1. Since the CD3-δ exon 1 begins 0.4 kb upstream of the *Bam*HI site (11), and assuming that the human CD3-γ exon 1 is similar in size to its mouse equivalent, the transcription initiation sites of the two genes are at most 2 kb apart. As in the case of the mouse genes, the exon localization provides sufficient evidence for a model in which transcription occurs in the opposite directions.

Based on comparisons of the nucleotide sequences of the human CD3-γ and -δ cDNAs, Krissansen *et al.* (19) reasoned that these genes had arisen after a gene duplication. The descriptions of the human and murine CD3-γ and -δ genes not only support that reasoning but add that this gene duplication occurred before human/mouse divergence. Although in terms of their exon/intron organization and sequence homology the CD3-γ and -δ genes are more different from the CD3-ε gene than from one another (S. Dunlap and H.C., unpublished data), all three CD3 genes are clearly related and must have arisen from a common ancestor. The divergence of the CD3-ε gene from the CD3-γ and -δ genes most likely occurred before CD3-γ diverged from CD3-δ. This may also indicate that their gene products play different roles in the signal transduction that occurs after the T-cell receptor triggering. Although CD3-γ and CD3-δ are very similar in their cytoplasmic regions, only CD3-γ becomes phosphorylated upon T-cell activation (4, 5).

Protein sequence analyses and secondary structure predictions suggest that the CD3 genes constitute a subgroup of the immunoglobulin gene superfamily (26). Previous work (13, 14) showed that the CD3-δ and -ε genes map close to each other on human chromosome 11 (band q23) and on mouse chromosome 9. The data presented here demonstrate that the three members of the CD3 gene family map to the same chromosome in humans and mice and that the CD3-γ gene is

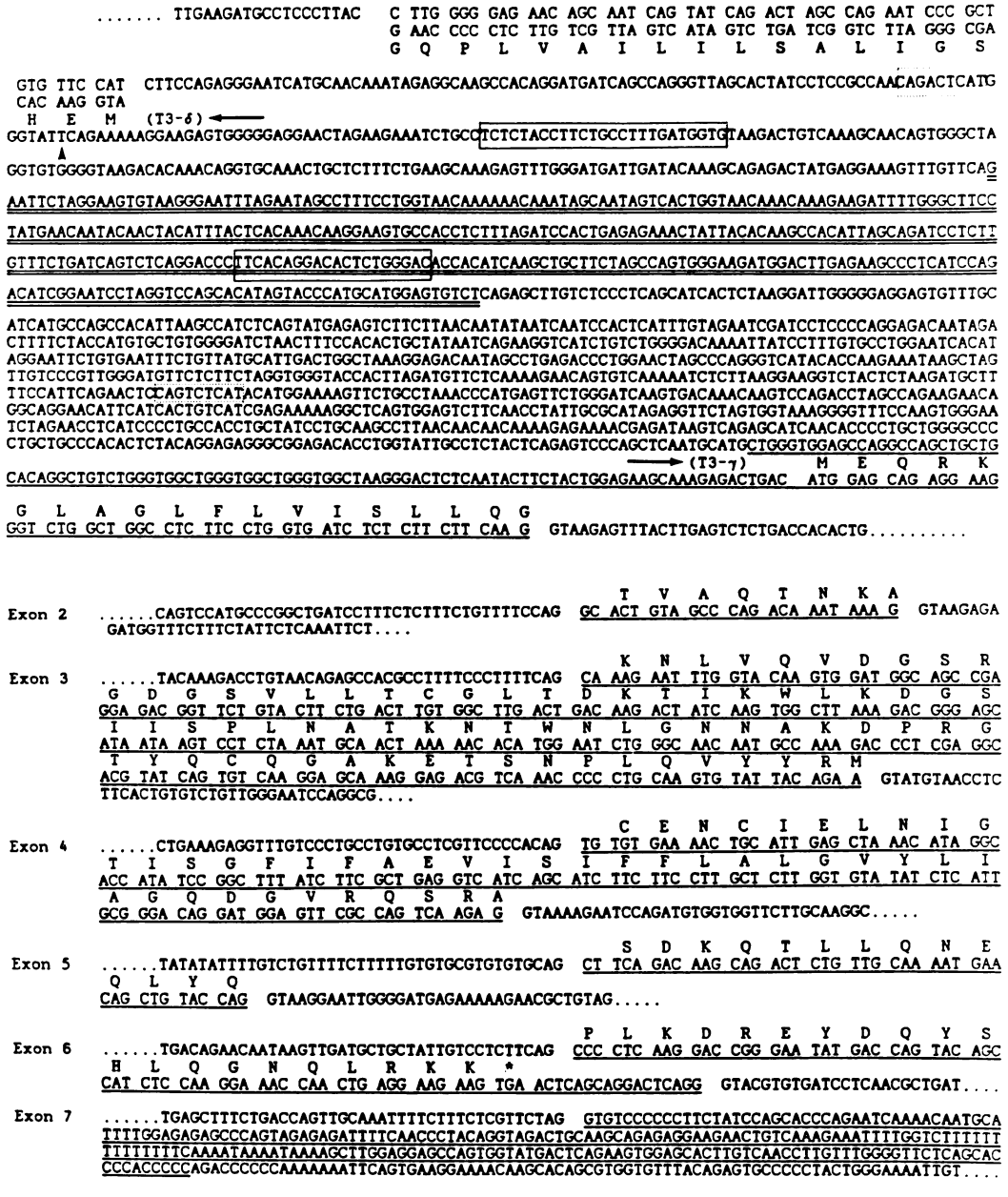

FIG. 2. Partial nucleotide sequence of the mouse CD3-γ and CD3-δ genes. The complete sequence of the DNA segment between the CD3-δ and CD3-γ (T3-δ and T3-γ) genes is presented. Since the transcription start site of the CD3-γ gene is not known, only the 5′ end of the cDNA pT134 (16) is indicated. Arrowhead shows the transcription initiation site in the CD3-δ gene. The segment underlined twice was sequenced two times using both λT3-γ1 and a *Eco*RI–*Acc* I fragment of the plasmid pTM-δ2 (11). The sequence above the twice-underlined sequence was published previously (11), and the amino acid sequence encoded by the first exon of the mouse CD3-δ gene is shown. The nucleotide sequences of the CD3-γ exons are underlined and are based upon the available cDNA sequences (16). The distances between the exons are given in Table 1. Predicted amino acid sequences are given in standard one-letter amino acid symbolism above the corresponding nucleotide sequences. Methods of subcloning and sequencing by the dideoxy chain-termination and chemical-modification techniques have been published (11, 15, 17).

Table 1. Comparison of the exon/intron organizations of the mouse CD3-γ and -δ genes

Exon		Size, bp		No. of amino acids encoded		Intron size, kb	
CD3-γ	CD3-δ	CD3-γ	CD3-δ	CD3-γ	CD3-δ	CD3-γ	CD3-δ
1	1	>159	153	18⅓	18⅓	≈5.6	≈3.0
2		24		8		0.780	
3	2	228	219	76	73	0.525	0.413
4	3	132	132	44	44	≈2.3	0.322
5	4	44	44	14⅔	14⅔	0.476	0.171
6	5	83	151	21	21	0.772	
7		>228		0			

located at 1.4 kb from CD3-δ. Preliminary results of pulsed-field gradient electrophoresis have shown that the CD3-γ and -δ genes are within 400 kb of the CD3-ε gene (H.C., unpublished data). More important than their close proximity is the indication that the CD3-γ and -δ genes are divergently transcribed. Since the two genes are expressed in all thymus-derived cells, including the earliest detectable cells in T-cell ontogeny, their transcription may be regulated coordinately. An analysis of the nucleotide sequence in the segment between the CD3-γ and -δ genes showed that both genes lack the "TATA" and "CAAT boxes" in their promoter regions. Two short segments of high homology were found in the promoter and 5′ flanking regions of the CD3-δ and -ε genes (H.C., unpublished data). Two of these regions are relatively close to the transcription start site of the human CD3-δ gene

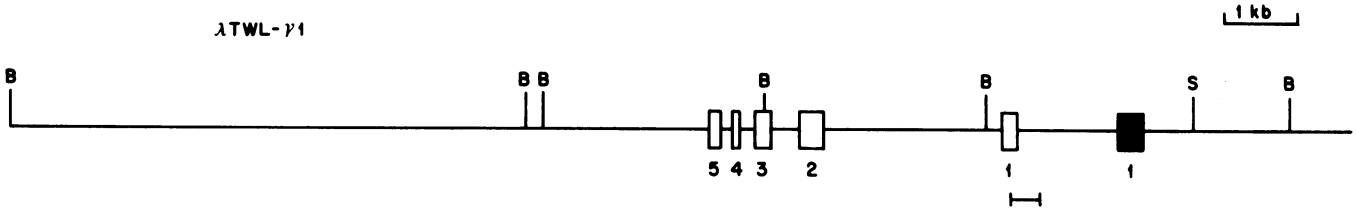

FIG. 3. Partial restriction map of the human CD3-δ and CD3-γ genes. A partial restriction map of the 20-kb insert of λTWL-γ1 was made to determine the distance between the first exons of the human CD3-δ and CD3-γ genes. The four probes used in this experiment are described in the text. The human genomic DNA probe covering an area that contains the transcription start site (−354 to +57) is indicated by a bar. Restriction sites: B, *Bam*HI; S, *Sal* I.

(positions −434 to −415 and −66 to −41) (H.C., unpublished data). If one or more of these regions play a role in transcriptional regulation, they may affect transcription of both genes. A very close linkage of genes that are divergently transcribed is a rare event (27–31) and in several of those instances it could be shown that regulatory elements exist in the area between the transcription start sites. In preliminary experiments, a tissue-specific enhancer element in the 3′ flanking region of the mouse CD3-δ gene was detected in an area where a DNase hypersensitive site was also found (K.G., unpublished data). Since enhancers can act over relatively long distances, this element might also affect the transcription of the CD3-γ gene, which is ≈6 kb from the enhancer area. The finding that the CD3-γ and the CD3-δ genes are in close proximity both in humans and in mice may therefore help in elucidating general principles of T-cell-specific gene expression.

We thank Drs. Yasuhiro Hashimoto and Alan Maxam for the mouse liver DNA library, Dr. Stuart Orkin for the human DNA library, Dr. Alan Maxam for helpful discussions, and John W. Lockhart for secretarial assistance. This research was supported by a grant from the Whitaker Health Sciences Fund (to H.S. and S.T.), an Investigator Award from the Cancer Research Institute (to H.S.), and by National Institutes of Health Grant AI17651 (to C.T.). K.G. is a Fellow of the Leukemia Society of America. H.C. is a Fellow of the Netherlands Cancer Fund (Koningin Wilhelmina Fonds). W.G.H. is a recipient of a postdoctoral fellowship from the National Institutes of Health (NS07723-02). T.L. was supported by a travel grant from the Leukemia Society of America.

1. Allison, J. P. & Lanier, L. L. (1987) *Annu. Rev. Immunol.* **5**, 503–540.
2. Toyonaga, B. & Mak, T. (1987) *Annu. Rev. Immunol.* **5**, 585–620.
3. Borst, J., Colligan, J. E., Oettgen, H., Pessano, S., Malin, R. & Terhorst, C. (1984) *Nature (London)* **312**, 455–458.
4. Samelson, L. E., Harford, J. B. & Klausner, R. D. (1985) *Cell* **43**, 223–231.
5. Oettgen, H., Pettey, C., Maloy, W. & Terhorst, C. (1986) *Nature (London)* **320**, 272–275.
6. Pettey, C., Alarcon, B., Malin, R., Weinberg, K. & Terhorst, C. (1987) *J. Biol. Chem.* **262**, 4854–4859.
7. Alarcon, B., Berkhout, B., Breitmeyer, J. & Terhorst, C. (1987) *J. Biol. Chem.*, in press.
8. Furley, A. J., Mitzutani, S., Weilbacher, K., Dhaliwal, H. S., Ford, A. M., Chan, L. C., Molgaard, H. V., Toyonaga, B., Mak, T., van den Elsen, P., Gold, D., Terhorst, C. & Greaves, M. F. (1986) *Cell* **46**, 75–87.
9. van Dongen, J. J. M., Quertermous, T., Bartram, C. R., Gold, D. P., Wolvers-Tettero, I. L. M., Comano-Bittes, W. M., Hooijkaas, H., Adriaansen, H. J., de Klein, A., Raghavachar, A., Ganser, A., Duhey, A. D., Seidman, J. G., van den Elsen, P. & Terhorst, C. (1986) *J. Immunol.* **47**, 365–370.
10. van Dongen, J. J. M., Krissansen, G. W., Wolvers-Tettero, I. L. M., Comans-Bitter, W., Adriaansen, H. J., Hooijkaas, H., van Wering, E. R. & Terhorst, C. (1987) *Blood*, in press.
11. van den Elsen, P., Georgopoulos, K., Shepley, B. A., Orkin, S. & Terhorst, C. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2944–2948.
12. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
13. van den Elsen, P., Bruns, G., Gerhard, D. S., Pravtcheva, D., Jones, C., Housman, D., Ruddle, F., Orkin, S. & Terhorst, C. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2920–2924.
14. Gold, D., van Dongen, J. J. M., Morton, C. C., Bruns, G. A. P., van den Elsen, P., van Kessel, A. H. M. G. & Terhorst, C. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1664–1668.
15. Koyama, T., Hall, L. R., Haser, W. G., Tonegawa, S. & Saito, H. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1609–1613.
16. Haser, W. G., Saito, H., Koyama, T. & Tonegawa, S. (1987) *J. Exp. Med.* **166**, 1186–1191.
17. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
18. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
19. Krissansen, G. W., Owen, M. J., Verbi, W. & Crumpton, M. J. (1986) *EMBO J.* **5**, 1799–1808.
20. Feinberg, A. P. & Vogelstein, B. (1983) *Ann. Biochem.* **132**, 6–13.
21. Saito, H., Kranz, D. M., Takagaki, Y., Hayday, A. C., Eisen, H. N. & Tonegawa, S. (1984) *Nature (London)* **309**, 757–762.
22. Saito, H., Kranz, D. M., Takagaki, Y., Hayday, A. C., Eisen, H. N. & Tonegawa, S. (1984) *Nature (London)* **312**, 36–40.
23. van den Elsen, P., Shepley, B.-A., Cho, M. & Terhorst, C. (1985) *Nature (London)* **314**, 542–544.
24. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
25. van den Elsen, P., Shepley, B.-A., Borst, J., Colligan, J., Markham, A. F., Orkin, S. & Terhorst, C. (1984) *Nature (London)* **312**, 413–418.
26. Gold, D., Clevers, H., Alarcon, B., Dunlap, S., Novotny, J., Williams, A. & Terhorst, C. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7649–7653.
27. Farnham, P. J. & Schimke, R. T. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3978–3982.
28. Mitchell, P. J., Carothers, A. M., Han, J. H., Harding, J. D., Kas, E., Venolia, L. & Chasin, L. A. (1986) *Mol. Cell. Biol.* **6**, 425–440.
29. Kay, R. J., Boissy, R. J., Russnak, R. H. & Candido, E. P. M. (1986) *Mol. Cell. Biol.* **6**, 3134–3143.
30. Goldsmith, M. R. & Kafatos, F. C. (1984) *Annu. Rev. Genet.* **18**, 443–487.
31. Karathanasis, S. K., Oettgen, P., Haddad, I. & Antonarakis, S. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8457–8461.