Volume 256, number 1,2, 185–191 FEB 07567 October 1989

The promoter regions of the T-cell receptor V9 γ (TRGV9) and V2 δ (TRDV2) genes display short direct repeats but no TATA box

Piona Dariavach and Marie-Paule Lefranc

Laboratoire d'Immunogénétique Moléculaire, URA CNRS 199, Université Montpellier II – Sciences et Techniques du Languedoc, Place E. Bataillon, 34060 Montpellier Cedex 1, France

Received 11 July 1989

T lymphocytes expressing the T-cell γδ receptor have been shown to express preferentially the T-cell receptor V9 γ(TRGV9) gene, in association with the T-cell receptor V2 δ(TRDV2) gene. In this paper, we report that the promoter regions of the TRDV2 and TRGV9 genes, which are preferentially expressed early in T-cell differentiation, display short direct repeats but no TATA box, in contrast to the Vγ genes belonging to subgroup I. The TCCTCAGT octanucleotide found 100 pb upstream of the ATG of the HD-Mar Vα transcript, a TcR Vα gene without a TATA box, is observed upstream of TRDV2 but not TRGV9. Of interest is the presence of a characteristic decanucleotide AGGTGGT(T)GAG in the promoter regions of both the TRDV2 and TRGV9 genes.

Promoter region; Receptor, γ T-cell; Receptor, δ T-cell; Gene, variable; Transcription; (Human)

1. INTRODUCTION

In human peripheral blood, about 3–5% of the circulating cells express the T-cell γ/δ receptor. The organization of the genes encoding the γ, δ chains and that of the corresponding loci have recently been elucidated. The human T-cell receptor γ (TRG) [1] locus which has been mapped on chromosome 7 [2] at band 7p15 [3], comprises two constant region genes (TRGC) [4,5] and 14 variable (TRGV) γ genes, belonging to four subgroups [6–8] (reviews [9–11]). Nine of them, five functional (V2, V3, V4, V5 and V8) and four pseudogenes (V1, V5P, V6, V7), belong to subgroup I, whereas subgroups II, III and IV each consist of a single gene, designated V9, V10 and V11, respectively [6–8]. Two pseudogenes, VA and VB, located upstream of V9 and V11, respectively, belong to none of these subgroups [7,8]. Three γ joining segments, JP1 [12–14], JP [16] and J1 [4], have been identified upstream of TRGC1, and two others, JP2 [12–14] and J2 [4] upstream of TRGC2. With its 14 variable genes spanning 100 kb, the two constant region genes and 5 joining segments covering less than 40 kb [15] and only 16 kb separating the most 3′ V gene (VII) from the most 5′ J (JP1) segment, the human TRG locus spans 160 kb of genomic DNA [16] (fig.1A).

The human T-cell receptor δ (TRD) locus is embedded in the T-cell receptor α locus between the Vα and Jα segments at 85 kb in 5′ from the Cα gene. It comprises a unique constant region gene

Correspondence address: M.-P. Lefranc, Laboratoire d'Immunogénétique Moléculaire, URA CNRS 199, Université Montpellier II – Sciences et Techniques du Languedoc, Place E. Bataillon, 34060 Montpellier Cedex 1, France

The nucleotide sequence data reported in this paper will appear in the EMBL, Genbank and DOBJ Nucleotide Sequence Databases under the accession numbers X15269-X15275.

Abbreviations: TRG, T cell receptor γ genes; TRGC, γ constant region genes (TRGC1 and TRGC2 for the two constant region genes); TRGJ, γ joining region gene segments (TRGJP1, JP, J1, JP2, J2 for the different J γ gene segments); TRGV, γ variable region genes (TRGV1, V2, V3, V4, V5, V5P, V6, V7, V8, VA, V9, V10, VB, V11 for the 14 variable genes in linkage order); TRD, T cell receptor δ genes; TRDC, δ constant region gene; TRDJ, δ joining region gene segments; TRDV, δ variable region genes; TcR, T cell receptor; kb, kilobases; bp, base pair.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50

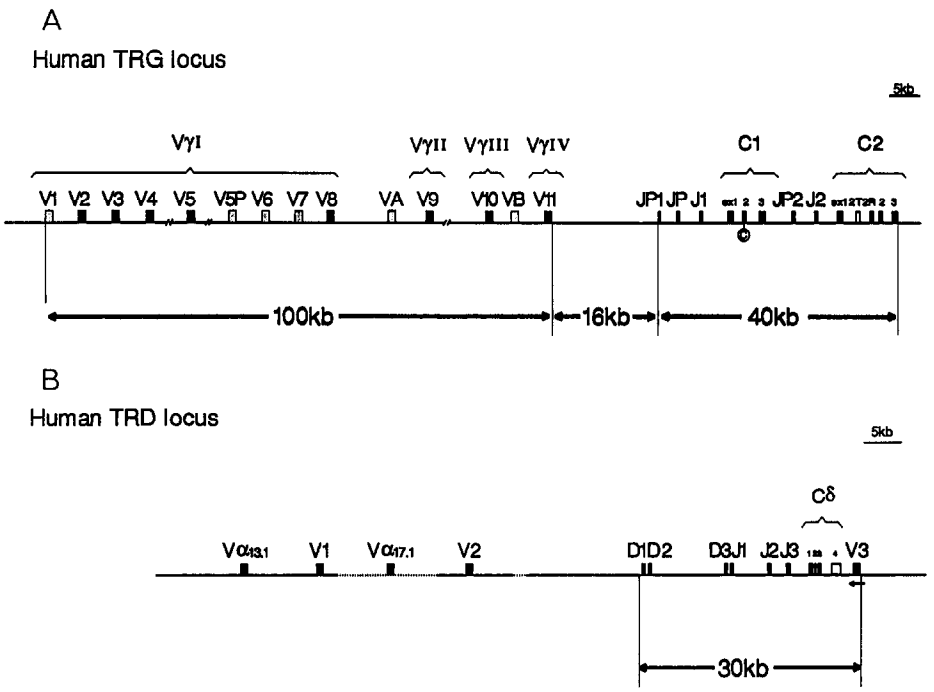

Fig.1. (A) Organization of the human T cell receptor γ TRG locus (from [16]). The human TRG locus comprises 14 TRGV genes, eight of which are functional and belong to four different subgroups [4, 6–8]. The functional V genes and the pseudogenes (see text) are represented by black and grey rectangles, respectively. The two TRGC genes [1,4,5] are shown with the associated TRGJ segments J1, J2 [4], JP [6], JP1 and JP2 [12–14]. In the TRGC2 gene, the dotted box represents the exon 2T characteristic of the allele C2(3X) [15]. (B) Schematic representation of the human T-cell receptor δ TRD locus [17–29]. The human TRD locus comprises a unique constant region gene TRDC preceded by the three TRDJ segments J1, J2 and J3 and three D segments D1, D2 and D3 [17–27]. V1 corresponds to the variable gene of the O-240/38 [17] and PEER [18] cDNA clones and V2 to the X13 [28] and LB117 [27] cDNA clones. V3 has recently been identified downstream of TRDC [26,27].

Cδ (TRDC) preceded by at least three joining gene segments (TRDJ) and three diversity segments [17–27]. The number of the variable δ genes (TRDV) seems limited. So far three of them V1 [17,18], V2 [27,28] and V3 [26,27] have been identified, the gene V3 lying in an inverted orientation, 3′ of the Cδ gene (fig.1B).

The majority of the human TcR γ/δ+ circulating cells express a receptor with a limited γ and δ chain combinatorial diversity that derived from the preferential TRGV9-JP and TRDV2-D-J1 rearrangement, respectively [30–32]. Here, we report the sequence of the promoter region of the TRDV2 gene, that of the TRGV9 gene, as well as those of the V γ genes belonging to subgroup I. We show that the promoter region of TRDV2 and TRGV9, in contrast to the V γ genes belonging to subgroup I, do not include the classical 'TATA' and 'CAAT' boxes but instead display short repeats in tandem, suggesting that the transcription of both TRDV2 and TRGV9 genes is initiated from a non-TATA promoter. A characteristic decanucleotide AGGTGGT(T)GAG is found in the promoter regions of both the TRDV2 and TRGV9 genes. This is of interest, since both TRDV2 and TRGV9 genes appear to be expressed early in T-cell differentiation [28,32].

2. MATERIALS AND METHODS

2.1. *Isolation of the 5′-TRDV2 nucleotide sequence*

A 600 bp fragment located upstream of the ATG codon of the TRDV2 gene has been sequenced. This fragment was isolated from the λLY67Vδ2 genomic clone, containing the TRDV2 gene [33].

2.2. *Isolation of the 5′-TRGV1, -V2, -V3, -V4, -V8 and -V9 nucleotide sequences*

The 5′ region of the V1, V2, V3, V4 and V8 genes (Vγ I subgroup genes) have been sequenced from the *Sac*I site located

236 bp upstream of the ATG codon [6]. The corresponding subclones were isolated respectively from λF1 and λF6 (V1), λSH4 (V2), λS1 and λSH4 (V3), λS13 (V4), λK20 (V8). These clones have previously been described [6]. A 2.7 kb *Hin*dIII-*Sac*I fragment containing the TRGV9 gene was isolated from the λ K20 clone [4].

2.3. *Subcloning and sequencing strategies*

Appropriate subclones were made in pUC vector [34]. Nucleotide sequence analysis was carried out by dideoxy chain-termination procedures [35] in M13 vectors [36], via exonuclease III-nuclease S_1 methods [37] or directed sequencing using known restriction enzyme sites.

3. RESULTS

The TRDV2 gene is preferentially expressed with the TRGV9 gene in T lymphocytes expressing the γ/δ receptor [28, 31, 32]. It was therefore of interest to analyse the promoter regions of both genes. The sequences of the region upstream of the ATG of TRDV2 isolated from λLY67Vδ2 [33] and that of TRGV9 isolated from λK20 [4, 7, 8] are shown in figs 2 and 3, respectively. The 5′ untranslated regions of TRDV2 and TRGV9 have been characterized to identify sequences which may have significance in the initiation of transcription. Although two A-T-rich regions are noted 500 and 580 bp, respectively, upstream of the initiation methionine codon ATG of TRGV9, it seems unlikely that they represent the conserved TATA box which has for consensus the conserved TATAAAA sequence [38]. Indeed, a 500 bp untranslated sequence would be very unusual, since a compilation of eukaryotic cellular mRNAs shows that 95% of them have 5′ untranslated regions less than 160 bp long [39]. We therefore looked for other sequences in the region more proximal to the V9 gene and in the V2 promoter region. No con-

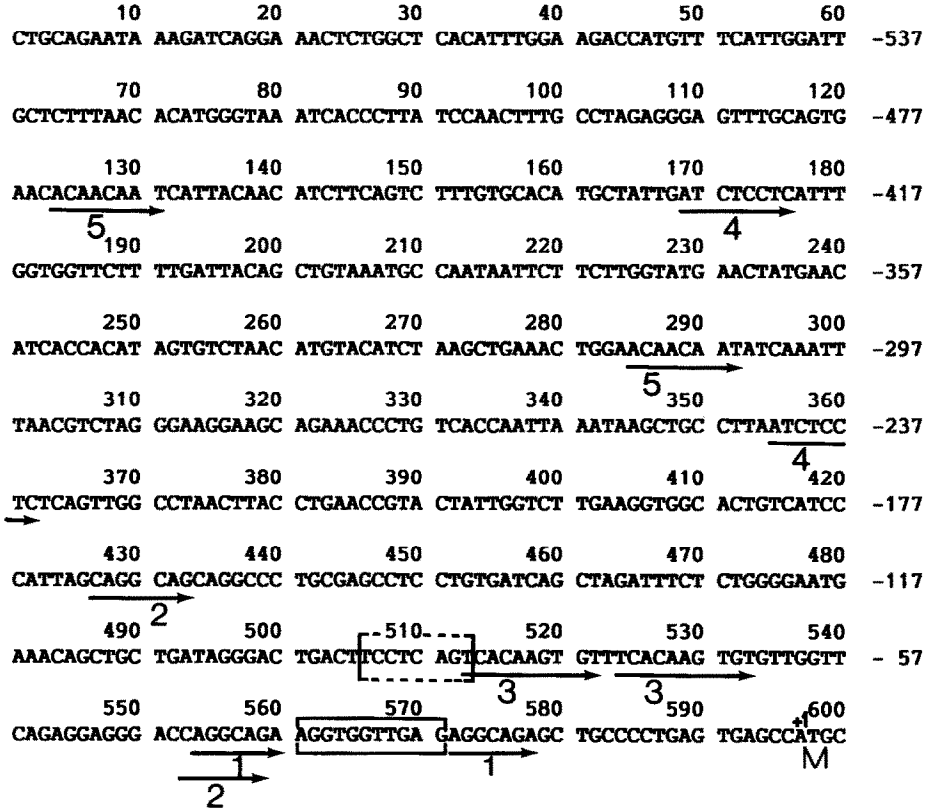

Fig.2. Nucleotide sequence of the 5′ untranslated region of the human TRDV2 gene. M indicates the initiation methionine codon (ATG). Pairs of direct repeats are shown by underlining arrows and are numbered. Nucleotide +1 denotes the A of the ATG encoding the initiation methionine codon of the TRDV2 gene and nucleotides preceding this codon are indicated by negative numbers on the right-hand side of the sequence. The TCCTCAGT octanucleotide, also found upstream of the HD-Mar Vα transcript, is shown by the dashed box. Except for the presence of an additional nucleotide, the boxed sequence is found to be identical to that in the promoter region of TRGV9 (see fig.3).

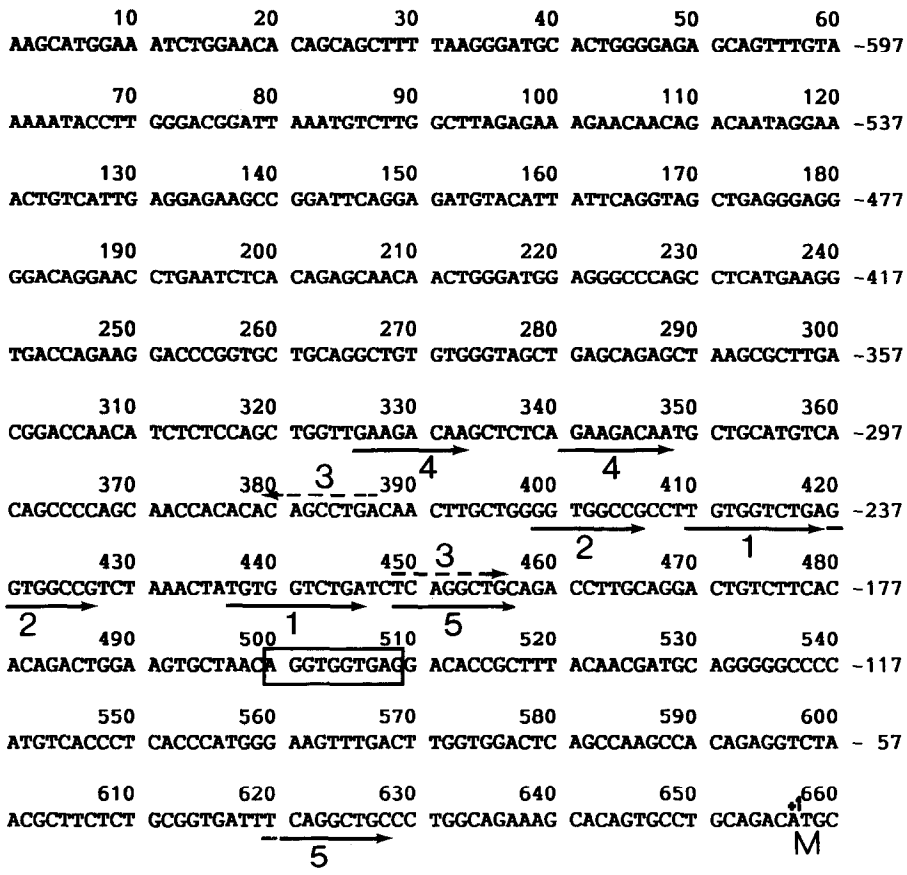

Fig.3. Nucleotide sequence of the 5′ untranslated region of the human TRGV9 gene. Only direct repeats equal to or greater than 8 nucleotides and less than 30 nucleotides distant are indicated by underlining arrows. Dashed arrows show inverted complementary sequences. The boxed sequence is found to be identical, except for the absence of one nucleotide (T), to that in the promoter region of the TRDV2 (see fig.2).

ventional TATA or CAAT [38] boxes could be found in either the TRGV9 or TRDV2 promoter regions. This is unlike most other eukaryotic promoters including promoters of human T-cell receptor β variable genes which do have these two sequences approx. 30 and 80 bp upstream of the transcription start sites [38]. Moreover, the TRDV2 and TRGV9 promoters do not possess the CCGCCC (or its inverted form GGGCGG) repeated sequences ('GC' box) which have been identified in promoters without typical TATA and CAAT boxes as, for example, promoters of 'house-keeping' genes [40, 41] or of viral genes [42]. However, these regions are rich in short 8–10 bp direct repeat sequences (underlining arrows in figs 2, 3). Two of them in tandem (repeats 1, 2) are localized between two complementary and reverse sequences in the TRGV9 promoter region (dashed arrows in fig.3). Note that the decanucleotide sequence of repeat 1 (TGTGGTCTGA) (fig.3) bears relaxed homology to an enhancer core sequence (5′-TGTGG(AAA/TTA)G-3′) that has been proposed on the basis of DNA sequence comparisons and in vitro mutagenesis experiments (review [43]).

Interestingly, despite the absence of homology between the promoter regions of TRDV2 and TRGV9, a characteristic decanucleotide AGGTGGT(T)GAG is found 36 and 157 bp upstream of the ATG codon of the TRDV2 and TRGV9 genes, respectively, with however an additional nucleotide (T) in the V2 sequence. The TRDV2 sequence also possesses the octanucleotide

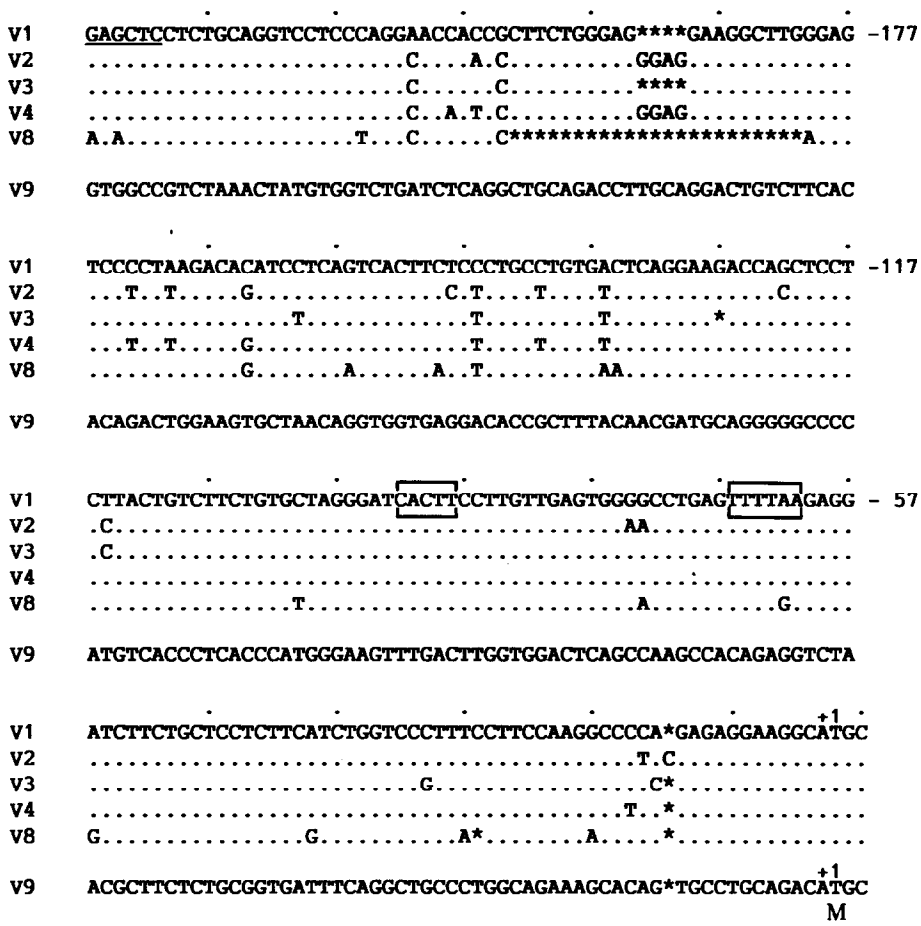

Fig.4. Nucleotide sequences of the 5′ untranslated region of the human Vγ I (V1, V2, V3, V4, V8) and Vγ II (V9) subgroup genes. Nucleotide +1 denotes the A of the ATG encoding the initiation methionine codon (M). The *Sac*I site is underlined in the TRGV1 sequence. The putative TATA and CACAT sequences are boxed in the TRGV1 sequence. Deletions are shown by asterisks. Dots indicate identical nucleotides. Since the 5′ region of the Vγ I subgroup genes and that of the V9 gene could not be aligned, these sequences are given in full.

TCCTCAGT which was found 100 bp upstream of the ATG of the HD-Mar Vα transcript [44], a Vα gene without a TATA or CAAT box. It is of interest that this octanucleotide occupies the same position in the promoter region of TRDV2. The sequencing of the promoter region of the TRGV1, V2, V3, V4 and V8 genes belonging to the Vγ I subgroup was carried out in order to search for the characteristic features of the TRDV2 and TRGV9 genes, i.e. the absence of the TATA and CAAT boxes and the presence of direct repeats and decanucleotide. In contrast to the TRGV9 gene, the promoter regions of the TRGV γ I genes shown in fig.4 seem to have conventional, although poorly conserved, TATA and CAAT promoters.

4. DISCUSSION

In the CD3+ TcR γ/δ+ lymphocytes, productive rearrangements involving the TRG1 region and the Vγ9 gene occur preferentially before rearrangements involving the TRG2 region during the process of T lymphocyte differentation [32]. In the CD3+ TcR α/β+ lymphocytes, the TRGV subgroup I genes are preferentially rearranged to the TRG2 region, but are rarely transcribed [45–48]. This observation suggests that the two TRG regions as well as the four TRGV gene subgroups are activated in a sequentially ordered fashion [32]. Thus, the TRGV9 gene which is transcribed in the majority of the CD3+ TcR

$\gamma/\delta+$ lymphocytes [31, 32] might be subject to cell-specific transcriptional control early in differentiation. Analysis of the TRDV2 and TRGV9 gene structure is therefore a crucial first step toward studying the mechanism which regulates their expression. Two features that have been reported in the promoter region of the CD3 δ gene [49, 50] are observed at the 5′-end of TRDV2 and TRGV9 (but not in the case of the V γ subgroup I genes): lack of TATA and CAAT boxes and presence of short repeats in tandem. These repeats may play a role in the activation of this gene, like that of a TATA box, by defining start positions for the RNA polymerase. The structure of these different promoter regions may be used by activation mechanisms specific of genes expressed during early T-cell differentiation, like the CD3 δ gene and the TRGV9 and TRDV2 genes. Note that other genes which are also expressed early in T-cell differentation, the CD3 γ [51] and ϵ [52] genes, the CD2 gene [53] and the mouse Thy-1 [54] lack TATA, CAAT and GC boxes. Finally, the presence of a characteristic decanucleotide in the promoter regions of the TRDV2 and TRGV9 genes may have some biological significance, since both TRDV2 and TRGV9 genes both appear to be expressed early in T-cell differentiation [31, 32].

Acknowledgements: We are indebted to F. Triebel and T. Hercend for the gift of the cDNA clone X13. We would like to thank G. Lefranc, T.H. Rabbitts and U. Schibler for helpful discussions. We are very grateful to A. Forster for his help in the sequencing of the 5′-region of the Vγ I subgroup genes. This work was supported by the Fondation pour la Recherche Médicale, the Association pour la Recherche sur le Cancer, the Centre National de la Recherche Scientifique, the Ligue Française contre le Cancer and the Ministère de la Recherche et de l'Enseignement Supérieur.

REFERENCES

[1] Lefranc, M.-P. and Rabbits, T.H. (1985) Nature 316, 464-466.
[2] Rabbitts, T.H., Lefranc, M.-P., Stinson, M.A., Sims, J.E., Schroder, J., Steinmetz, M., Spurr, N.L., Solomon E. and Goodfellow, P.N. (1985) EMBO J. 4, 1461-1465.
[3] Murre, C., Waldmann, R.A., Morton, C.C., Bongiovanni, K.F., Waldmann, T.A., Shows, T.B. and Seidman, J.G. (1985) Nature 316, 549-552.
[4] Lefranc, M.-P., Forster, A. and Rabbits, T.H. (1986) Nature 319, 420-422.
[5] Lefranc, M.-P., Forster, A. and Rabbits, T.H. (1986) Proc. Natl. Acad. Sci. USA 83, 9596-9600.
[6] Lefranc, M.-P., Forster, A., Baer, R., Stinson, M.A. and Rabbitts, T.H. (1986) Cell 45, 237-246.
[7] Forster, A., Huck, S., Ghanem, N., Lefranc, M.-P. and Rabbitts, T.H. (1987) EMBO J. 6, 1945-1950.
[8] Huck, S., Dariavach, P. and Lefranc, M.P. (1988) EMBO J. 7, 719-726.
[9] Lefranc, M.-P., Foster, A. and Rabbitts T.H. (1987) in: The T Cell Receptor (Kappler, J. and Davis, M. eds), vol. 73, pp. 25-29, A.R. Liss , New York.
[10] Lefranc, M.-P. (1988) Biochimie 70, 901-908.
[11] Lefranc, M.-P. and Rabbitts, T.H. (1989) Trends Biochem. Sci, in press.
[12] Huck, S. and Lefranc, M.-P. (1987) FEBS Lett. 224, 291-296.
[13] Quertermous, T., Strauss, W.M., Van Dongen, J.J.M. and Seidman, J.G. (1987) J. Immunol. 138, 2687-2690.
[14] Tighe, L., Forster, A., Clark, D., Boylston, A., Lavenir, I. and Rabbitts, T.H. (1987) Eur. J. Immunol. 17, 1729-1736.
[15] Buresi, C., Ghanem, N., Huck, S., Lefranc, G. and Lefranc, M.-P. (1989) Immunogenetics 29, 161-172.
[16] Lefranc, M.-P., Chuchana, P., Dariavach, P., Nguyen, C., Huck, S., Brockly, F., Jordan, B. and Lefranc, G. (1989) Eur. J. Immunol., in press.
[17] Hata, S., Brenner, B. and Krangel, M.S. (1987) Science 238, 678-682.
[18] Loh, E.Y., Lanier, L.L., Turck, C.W., Littman, D.R., Davis, M.M., Chien, Y.-H. and Weiss, A. (1987) Nature 330, 569-572.
[19] Boehm, T., Baer, R., Lavenir, I., Forster, A., Waters, J.J., Nacheva, E. and Rabbitts, T.H. (1988) EMBO J. 7, 385-394.
[20] Hata, S., Satyanarayana, K., Devlin, P., Band, H., McLean, J., Strominger, J.L., Brenner, M.B. and Krangel, M.S. (1988) Science 240, 1541-1544.
[21] Isobe, M., Russo, G., Haluska, F.G. and Croce, C.M. (1988) Proc. Natl. Acad. Sci., USA 85, 3933-3937.
[22] Baer, R., Boehm, T., Yssel, H., Spits, H. and Rabbitts, T.H. (1988) EMBO J. 7, 1661-1668.
[23] Griesser, H., Champagne, E., Tkachuk, D., Takihara, Y., Lalande, M., Baillie, E., Minden, M. and Mak, T.W. (1988) Eur. J. Immunol. 18, 641-644.
[24] Takihara, Y., Tkachuk, D., Michalopoulos, E., Champagne, E., Reimann, J., Minden, M. and Mak, T.W. (1988) Proc. Natl. Acad. Sci., USA 85, 6097-6101.
[25] Hockett, R.D., De Villartay, J.-P., Pollock, K., Poplack, D. G., Cohen, D.I. and Korsmeyer, S.J. (1988) Proc. Natl. Acad. Sci. USA 85, 9694-9698.
[26] Loh, E.Y., Cwila, S., Serafini, A.T., Philipps, J.H. and Lanier, L.L. (1988) Proc. Natl. Acad. Sci., USA 85, 9714-9718.
[27] Hata, S., Clabby, M., Devlin, P., Spits, H., De Vries, J.E. and Krangel, M.S. (1989) J. Exp. Med. 169, 41-57.
[28] Triebel, F., Faure, F., Mami-Chouaib, Jitsukawa, S., Griscelli, A., Genevee, C., Roman-Roman, S. and Hercend, T. (1988) Eur. J. Immunol. 18, 2021-2027.
[29] Sturm, E., Braackman, E., Bontrop, R.B., Chuchana, P., Van de Griend, R.J., Koning, F., Lefranc, M.-P. and Bolhuis, L.H. (1989) Eur. J. Immunol., in press.

[30] Jitsukawa, S., Faure, F., Lipinski, M., Triebel F. and Hercend T. (1987) J. Exp. Med. 166, 1192–1197.
[31] Triebel, F., Faure, F., Graziani, M., Jitsukawa, S., Lefranc, M.-P. and Hercend, T. (1988) J. Exp. Med. 167, 694–699.
[32] Triebel, F., Lefranc, M.-P. and Hercend, T. (1988) Eur. J. Immunol. 18, 789–794.
[33] Dariavach, P. and Lefranc, M.-P. (1989) Nucleic. Acids Res., submitted.
[34] Vieira, J. and Messing, J. (1982) Gene 19, 259–268.
[35] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci., USA 74, 5463–5467.
[36] Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. 143, 161–178.
[37] Guo, L.H., Yang, R.C.A. and Wu, R. (1983) Nucleic Acids Res. 11, 5521–5540.
[38] Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349–383.
[39] Kozak, M. (1984) Nucleic Acids Res. 12, 857–872.
[40] Reynolds, G.A., Basu, S.K., Osborne, T.F., Chin, D.J., Gel, G., Brown, M.S., Goldstein, J.E. and Luskey, K.L. (1984) Cell 38, 275–285.
[41] Valerio, D., Duyvesteyn, M.G., Dekker, B.M., Weeda, G., Berkvens, T.M., Van der Voom, L., Van Ormondt, H. and Van der Eb, A. (1985) EMBO J. 4, 437–443.
[42] Fromm, M. and Berg, P.J. (1986) Mol. Appl. Genet. 1, 457–481.
[43] Sassone-Corsi, P. and Borelli, E. (1986) Trends Genet. 2, 215–219.
[44] Luria, S., Gross, G., Horowitz, M. and Givol, D. (1987) EMBO J. 6, 3307–3312.
[45] Boehm, T.L., Werle, A., Ganser, B., Korhuber, B. and Drahovsky, D. (1987) Eur. J. Immunol. 17, 1593–1597.
[46] Le Paslier, D., Chen, Z., Loiseau, P., Cohen, D. and Sigaux, F. (1987) Blood 70, 637–646.
[47] Migone, N., Casarotti, G., Francia di Celle, P., Lusso, P., Foa, R. and Lefranc, M.-P. (1988) Eur. J. Immunol. 18, 173–178.
[48] Moisan, J.-P., Bonneville, M., Bouyge, I., Moreau, J.-F., Soulillou, J.-P. and Lefranc, M.-P. (1989) Hum. Immunol. 24, 95–110.
[49] Tunnacliffe, A., Sims, J.E. and Rabbitts, T.H. (1986) EMBO J. 5, 1245–1252.
[50] Van den Elsen, P., Georgopoulos, K., Shepley, B.A., Orkin, S. and Terhorst, C. (1986) Proc. Natl. Acad. Sci. USA 83, 2944–2948.
[51] Saito, H., Koyama, T., Georgopoulos, K., Clevers, H., Haser, W.G., LeBien, T., Tonegawa, S. and Terhorst, C. (1987) Proc. Natl. Acad. Sci. USA 84, 9131–9134.
[52] Clevers, H.C., Dunlap, P., Wileman, T.E. and Terhorst, C. (1988) Proc. Natl. Acad. Sci. USA 85, 8156–8160.
[53] Diamond, D.J., Clayton, L.K., Sayre, P.H. and Reinhertz, E. (1988) Proc. Natl. Acad. Sci. USA 85, 1615–1619.
[54] Giguere, V., Isobe, K.I, and Grosfeld, F. (1985) EMBO J. 4, 2017–2024.