Two inflammatory mediator cytokine genes are closely linked and variably amplified on chromosome 17q

Steven G.Irving+, Peter F.Zipfel[1,5], Jeff Balke[2,5], O.Wesley McBride[3], Cynthia C.Morton[4], Parris R.Burd[5], Ulrich Siebenlist[5] and Kathleen Kelly*
Laboratory of Pathology, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA, [1]Bernhard Nocht Institute for Tropical Medicine, Bernhard Nocht Strasse 74, D-2000 Hamburg 4, FRG, [2]Department of Internal Medicine, University of Minnesota Hospital and Clinic, Harvard St. and E.River Rd, Minneapolis, MN 55455, [3]Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, [4]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, 75 Francis St, Boston, MA 02115 and [5]Laboratory of Immunoregulation, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA

Received February 23, 1990; Revised and Accepted May 4, 1990 EMBL accession no. X52149

ABSTRACT

Mitogenic stimulation of resting T cells results in the *de novo* transcription of a large number of genes including those encoding regulatory molecules such as lymphokines. The genomic organization of two newly described induced lymphokine genes, 464.1 and 744.1, has been determined. 464.1 and 744.1 appear to be the human homologues of the recently cloned murine macrophage inflammatory proteins, MIP-1α and MIP-1β, respectively. The 464.1 and 744.1 genes share 55% amino acid homology and demonstrate parallel regulation of induced expression in T cells. It was therefore of interest to observe that these genes are closely linked in the human genome, separated by 14 kb, and are organized in a head to head fashion. Each of the genes is present in an additional nonallelic copy (referred to as 464.2 and 744.2) as part of an apparent amplification unit in the genome of many individuals. The 464.2 gene is expressed and potentially encodes a protein highly related to 464.1, varying in 5 of 92 amino acids. As expected, 464.2 and 744.2 are also closely linked to each other as determined by population linkage disequilibrium studies. Individuals bearing a chromosome with a third amplification event, involving a 464-related gene but not a 744-related gene, are also infrequently observed. These genes are all located on chromosome 17 in bands q11-q21, the region implicated in von Recklinghausen neurofibromatosis (NF1) and in acute promyelocytic leukemia (AML-M3).

INTRODUCTION

The receipt of a mitogenic stimulus by a quiescent T cell sets into motion a series of biochemical and transcriptional events ultimately leading to cellular proliferation and immunologic function (1). Recently we have demonstrated the mitogen induced expression of two lymphokine genes whose mRNA is regulated similar to the IL-2 gene (2,3). The two genes, 464.1 and 744.1 (previously called pAT 464 and pAT 744; ref. (3)), share sequence similarity (45% at the nucleotide level and 55% at the amino acid level) and demonstrate coordinate regulation of mRNA expression. Comparison of the amino acid sequences of 464.1 and 744.1 with published lymphokine sequences demonstrated that they belong to a family of low molecular weight secreted proteins (4). Based on nucleic acid homology comparisons, the T cell products which we have described appear to be the human homologues of the recently identified macrophage induced protein (MIP-1) genes, MIP-1α (464.1) and MIP-1β (744.1), two LPS-induced monocyte-derived murine factors (5–7). These cytokines are most likely direct effectors of inflammatory responses as subcutaneous injection of MIP-1 (a mixture of the two peptides) into mouse footpads causes a local inflammatory response, primarily characterized by the infiltration of neutrophils (5), and have activity as an endogenous pyrogen which acts by a mechanism independent of TNF and IL-1 (8). *In vitro*, MIP-1 is weakly chemokinetic for neutrophils and triggers a respiratory burst in these cells (5).

In this paper, we describe the genomic organization of 464.1 and 744.1 and their chromosomal location to the long arm of chromosome 17 at q11-q21. We have found that these evolutionarily related genes are closely linked in the human

* To whom correspondence should be addressed

+ Present address: Department of Pathology, Georgetown University School of Medicine, Washington, DC 20007, USA

genome, and we present evidence for chromosomal haplotypes containing different numbers of these genes in the human population.

MATERIALS AND METHODS

Isolation of cDNA and genomic clones

pAT 464.2 cDNA was isolated as described previously for pAT 464.1 (3). Genomic clones were isolated from a library prepared by partial Mbo I digestion of human leukocyte DNA and cloning into the Bam HI site of the EMBL-3 vector.

DNA Sequence Analysis

cDNA and genomic inserts were cloned into Bluescribe M13+ and were sequenced by the dideoxy chain termination method (9) using Sequenase II (USB) and both universal and custom oligonucleotide primers. All data shown have been derived from overlapping sequence determined on both strands of the clones. The nucleotide sequence data for the 464.2 cDNA will appear in the EMBL Data Library under the accession number X52149.

Polymerase Chain Reaction

Limited stretches of the genomic sequence 5′ of the coding sequence of 464.1 and 464.2 were amplified by the polymerase chain reaction (PCR) method using the GeneAmp kit (Cetus). Amplified DNA was purified on a 6% acrylamide gel.

Probe preparation and hybridization

cDNA probes were labelled with ^{32}P-dCTP by the random primer method (10). PCR probes were labelled as follows: PCR primers (2.5 ug each) were annealed to the amplified segment (50 ng) by boiling for 2 min followed by slow cooling (approx. 1 h) to room temperature. The template-primer mix was then adjusted to the reaction conditions used for random primer labelling (10) except that no additional primers were added.

Detection of DNA restriction fragment length polymorphisms (RFLPs)

DNA was isolated from the peripheral leukocytes of 39 unrelated normal individuals, digested with restriction endonucleases (Eco RI, Hind III, Bam HI, Xba I, Sst I, Taq I, Pvu II, Pst I, Bgl II, Msp I, Eco RV and Kpn I) and subjected to Southern blotting. The distribution of genotypes at each polymorphic restriction site was readily determined from the Southern blots. The frequency of each allele was then determined by standard methods. Allele frequencies at the 464.2 and 744.2 loci are approximations based upon a simplifying assumption that the presence of only one allelic band was considered to represent homozygousity for that allele. Hence it is likely that some individuals classified as homozygotes were actually hemizygous for these alleles. The number of heterozygotes would be decreased but this was not detected when the number of observed heterozygotes in the population was compared to that predicted by Hardy-Weinberg equilibrium. This suggests that the assumption did not create a serious error in these determinations.

Nonrandom association between RFLPs

Tests for linkage disequilibrium were performed for each pair of RFLPs and quantified by the standardized disequilibrium statistic , where

$$\phi = (n_1)(n_4)-(n_2)(n_3)/[(n_1+n_2)(n_3+n_4)(n_1+n_3)(n_2+n_4)]^{1/2}$$

and n_1, n_2, n_3 and n_4 are the numbers of $++$, $+-$, $-+$ and $--$ haplotypes respectively (11,12) for the presence (+) or absence (−) of the restriction site at each locus. The statistical significance of ϕ in a sample size of n chromosomes was tested from the value $n\phi^2$ which is asymptotically distributed as a χ^2 random variable with 1 df (12,13).

Evolutionary distances

The evolutionary distance between a pair of sequences was measured by number of nucleotide substitutions per site using the formula of Jukes and Cantor (14):

$$\delta = -(3/4)\ln[1-(4/3)\pi],$$

where π is the proportion of nucleotides that are different between the two sequences under consideration. The δ value was obtained for each of the three nucleotide positions of codons and the phylogenetic relationships were determined by averaging the δ's for either the first and second positions or for all three nucleotide positions. The average of all three nucleotide positions is presented here although the evolutionary distances measured by either method were identical for all the genes examined in this study. This last point suggests an absence of selective pressure on much of the primary amino acid sequence as third position mutation rates for many proteins differ from first and second position mutation rates as third position changes are generally selectively neutral and occur at a uniform rate (even for proteins whose rate of amino acid-altering substitution varies ten-fold (15)).

Somatic cell hybrids

Panels of human-hamster and human-mouse somatic cell hybrids containing subsets of human chromosomes in a rodent chromosome background were prepared and analyzed as described (16−18). The human chromosome content of each hybrid was determined by isoenzyme analysis, Southern analysis with chromosomally localized probes, and karyotyping in some cases. DNAs were isolated from the same expanded cell populations used for characterization of human chromosome content. These DNAs were digested with Eco RI, electrophoresed in 0.7% agarose gels, and blotted onto nylon membranes.

In situ hybridization

In situ hybridization of a 3.5 kb Sst I-Sst I fragment contained in genomic clone λ464.1 and which included the 744.1 gene (shown in Fig. 2) was performed as described (19).

RESULTS

Isolation of two distinct 464-related cDNA clones

In the course of screening cDNA libraries, made from mitogen activated human T cell total poly A$^+$ RNA, with a partial pAT 464.1 cDNA insert, we isolated multiple pAT 464-hybridizing clones. The DNA sequence of one clone (464.2) revealed a number of differences relative to other full length 464 clones, represented here by 464.1 (Fig. 1A). These differences included two insertions (one nucleotide and two nucleotides respectively, separated by one nucleotide) resulting in an altered amino acid sequence from phe (residue 21 of 464.1) to val-leu (residues 21−22 of 464.2) as well as three additional single base substitutions resulting in ser to pro at 464.2 residue 25, ser to gly at 464.2 residue 62 and gly to ser at 464.2 residue 70. The remaining five base changes are silent with regard to the encoded

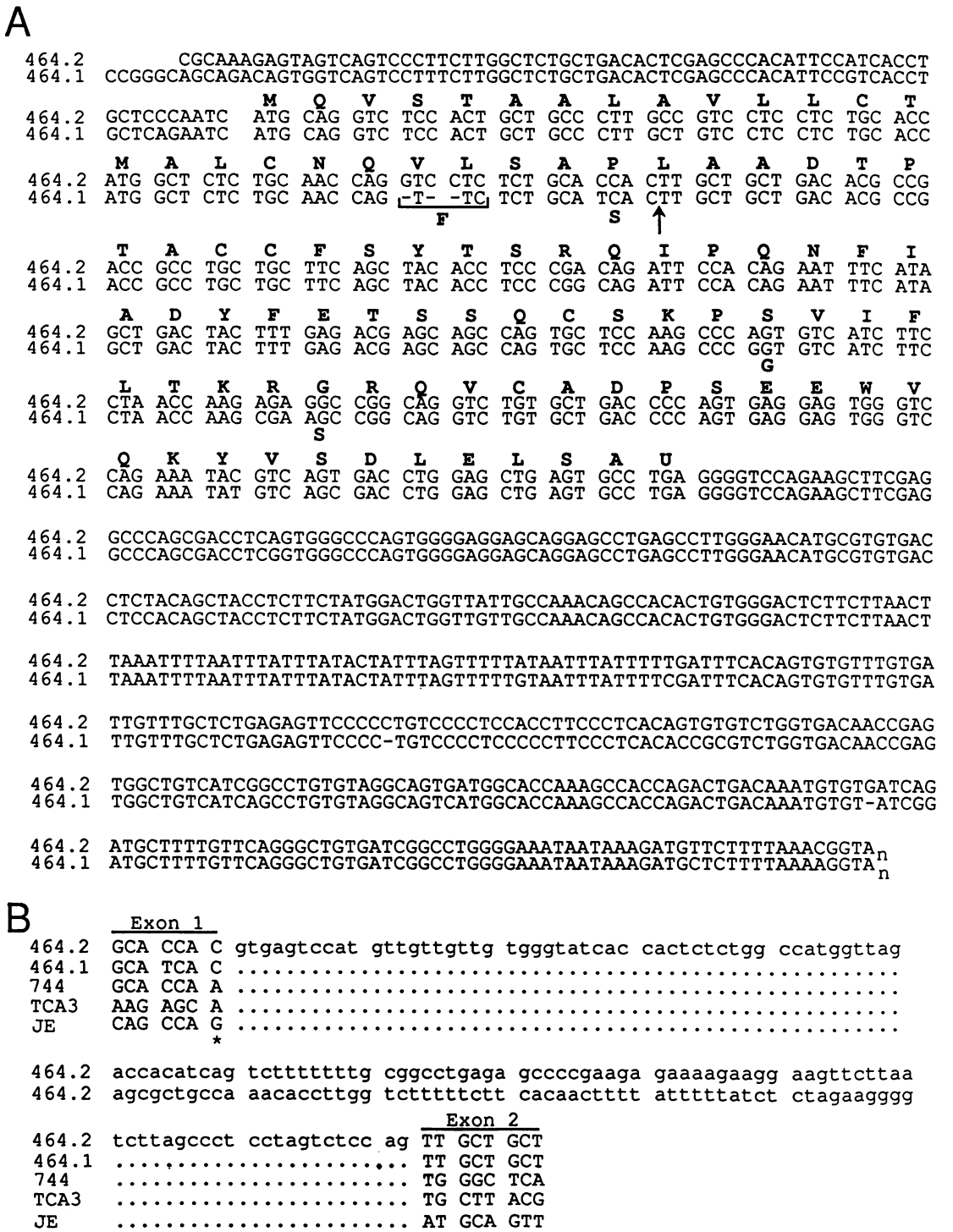

Figure 1. *Nucleotide sequence of 464.1 and 464.2 cDNAs.* Panel A shows the nucleotide sequence for 464.1 and 464.2 and the predicted amino acid sequence for 464.2. Amino acid differences in 464.1 (relative to 464.2) are shown under the 464.1 nucleotide sequence. The arrow shows the boundary between exons 1 and 2. Panel B shows the 5′ end of intron 1 and adjacent exon 1 and 2 sequences. The intron sequence retained in the 464.2 cDNA isolate is shown in lower case letters; the trinucleotide ...cag... at the 3′ end of the intron sequence is part of the cryptic splice donor site (...cag/gtatga...) found in the genomic clone. The corresponding splice junctions for four other related genes, 464.1, 744.1, JE and TCA3 are also shown; the * denotes the −1 position.

amino acids. An additional 192 nucleotides in pAT 464.2 (Fig. 1B) relative to pAT 464.1 appear to derive from unspliced intron 1 sequences (see below).

Isolation of genomic clones containing 744.1, 464.1 and 464.2 loci

To characterize the structure and genomic organization of 744- and 464-related genes, hybridizing clones were isolated from a human genomic library. Surprisingly, we discovered that a single genomic clone, λ464.1 contained both pAT 464- and pAT 744-hybridizing sequences. Restriction analysis, coupled with selective sequencing established that clone λ464.1 contained the 464.1 gene at a distance of approximately 14 kb from 744 with the direction of transcription for the two genes arranged in a head

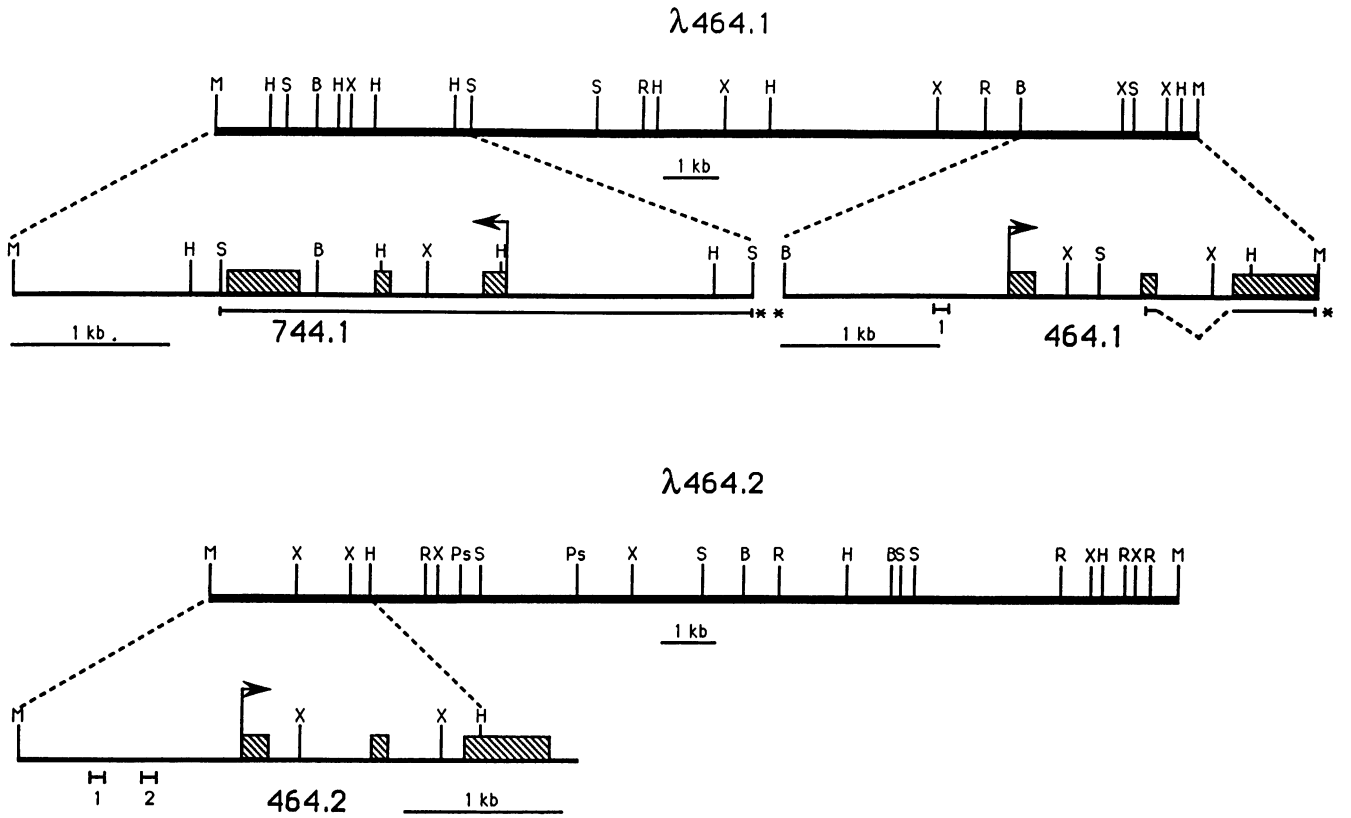

Figure 2. *Genomic organization of 464.1, 464.2 and 744.1.* The inserts of the two genomic clones λ464.1 and λ464.2 are represented as heavy lines; shown are the Bam HI (B), Hind III (H), Eco RI (R), Sst I (S) and Xba I (X) restriction sites for the entire clone as well as the Mbo I (M) cloning site flanking each insert; Pst I (Ps) sites were determined for the leftmost 10 kb of λ464.2 and the rightmost 3.2 kb of λ464.1. The regions in which the cDNAs have been mapped are expanded and the direction of transcription for each gene is shown. The 3 exons comprising each gene are shown as shaded boxes. The numbered brackets indicate the locations of the PCR-generated probes 1 and 2 (used in the experiments presented in Fig. 3C). The cDNA probe used in Figs. 3A and 3B is marked *. The genomic probe used for the *in situ* hybridization (Fig. 4) is marked **.

to head fashion (Fig. 2). In addition, a second clone, λ464.2, appeared to overlap with λ464.1 by approximately 3 kb in the region containing the 464.1 gene. Sequence analysis however, revealed that λ464.2 encoded the 464.2 cDNA and additional sequence information obtained from the λ464.1 and λ464.2 genomic clones has revealed the following. The exon boundaries for 464.1 as deduced from the sequence of a full length cDNA clone and the λ464.1 genomic clone are shown in Fig. 1B. A comparison of 464.1 exon boundaries to those established for the related genes, JE (20) or TCA3 (21,22) reveals perfect conservation of exon/intron boundaries relative to the encoded proteins aligned to obtain maximal homology (Fig. 1B). By comparison to the organization of the 464.1 gene, it was determined that 192 unique nucleotides in the pAT 464.2 cDNA clone were located at the presumed junction of exons 1 and 2 and thus may have derived from intron 1 as a result of an unexpected splicing event. Sequencing the relevant regions of the genomic clone λ464.2 revealed that the 192 nucleotide segment found in the pAT 464.2 cDNA appeared to originate from a splice between a cryptic donor (...CAG/GTATGA...) at nucleotide 193 of intron 1 and the normal splice acceptor at the intron 1-exon 2 boundary (Fig. 1B). While the normal donor sequence at the 3′ end of exon 1 in pAT 464.2 adhered to the GT-AG rule and generally conformed to the consensus of splice junction sequences (23), deviation from the consensus sequence for splice donors was observed in that a C rather than the more common G was found at the −1 position. Interestingly, although only 6% of the donor sequences catalogued in (23) had a C at the −1 position, it can be seen that a number of other closely related genes (i.e. 464.1 and 744.1, TCA3) also lack the more commonly found G (present in 73% of the donor sequences examined) at this position (Fig. 1B). Although the pAT 464.2 cDNA isolated by us did contain intron sequences which if translated directly would introduce multiple stop codons in all 3 reading frames, we believe that this clone is unusual and does not imply a functionally inactive gene for the following reasons. The 464.2 gene contains acceptable splice donor and acceptor sites and, more importantly, these sequences are evolutionarily conserved and characteristic of other family members. In addition, we have not observed aberrant mRNA species on Northern blots.

Comparison of 464- and 744- hybridizing genomic band patterns in multiple individuals

The differences noted in the nucleotide sequences between 464.1 and 464.2 could represent extensive polymorphism between two alleles of the 464 gene or, conversely, two highly related loci resulting from duplication in the human genome. To resolve this point, we performed Southern blot analyses of DNA from 39 unrelated individuals, cut with various restriction enzymes. In the panel of DNA samples examined, two kinds of variability were observed. Simple two allele restriction fragment length polymorphisms, due to the presence or absence of a restriction site, were found in Pst I (Fig. 3B) and Sst I (data not shown) digests. In addition, there was a systematic variation in the total number of hybridizing restriction fragments among different

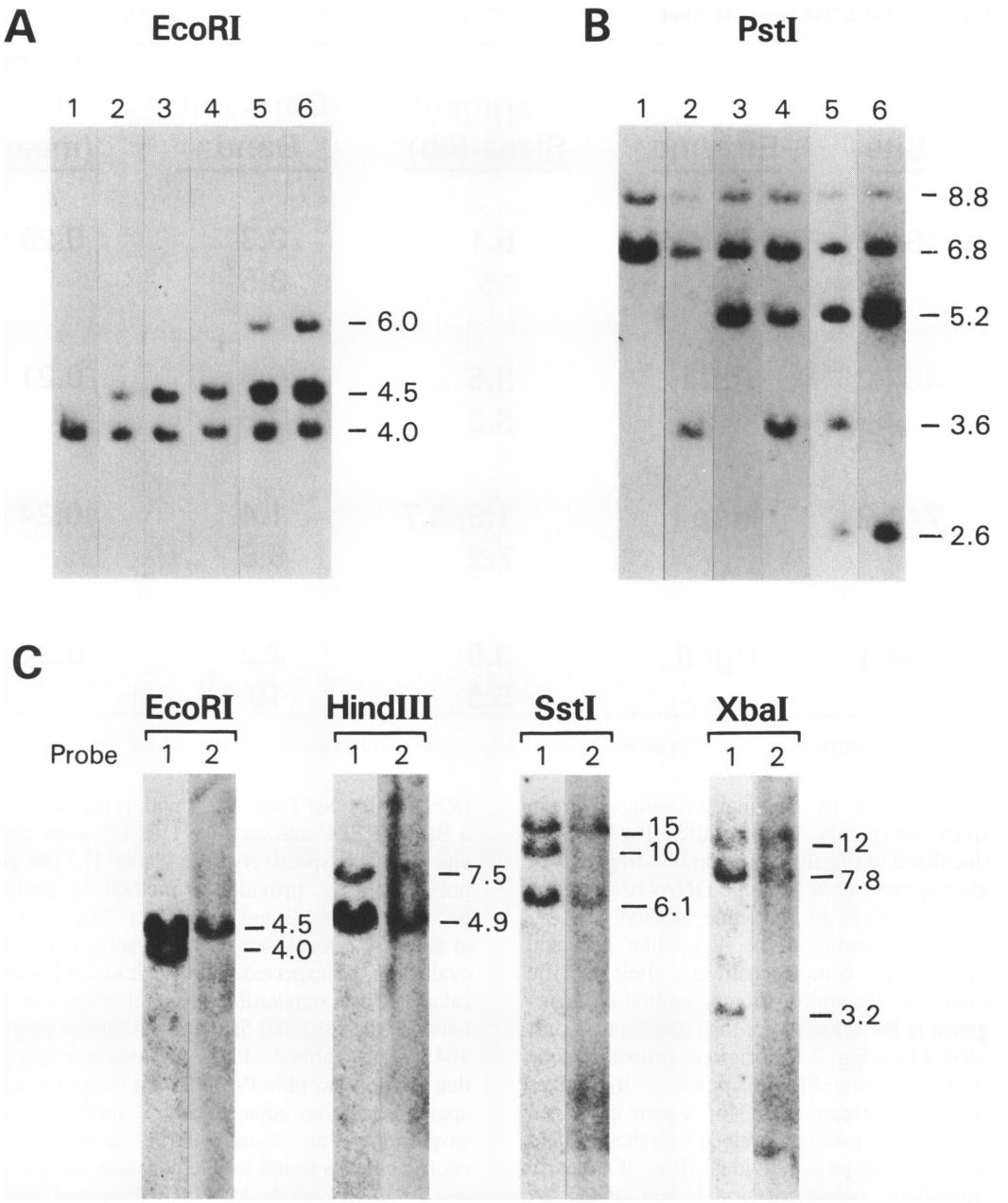

Figure 3. *Southern blot analysis of 464-related genes.* Panels A and B show the hybridization patterns obtained with a 464.1 cDNA probe (shown in Fig. 2) against blots of Eco RI and Pst I digested DNA from six normal unrelated individuals; the same DNAs are present in both panels. The 4.0 kb, 4.5 kb and 6.0 kb bands (A) are nonallelic fragments which represent polymorphism for copy number. The 4.0 kb band (A) and the 6.8 kb band (B) were found in all individuals examined. The 4.5 kb band (A) and the 5.2 kb plus 3.6 kb bands (B) are absent in lane 1 and the 6.0 kb (A) and 2.6 kb (B) bands are present only in lanes 5 and 6. The 5.2 kb and 3.6 kb bands represent a conventional two allele RFLP due to the presence or absence of a Pst I restriction site. This RFLP can be assigned to the reduplicated 464 gene (i.e. 464.2) since neither allele is found in lane 1. In panel C the hybridization patterns obtained by using probes of limited extent are shown. Probe 1 is common to 464.1 and 464.2 while probe 2 is specific for 464.2. (Probe 2 contains a sequence related to the Alu I repetitive element; these blots were washed at 62°C in 0.1× SSC to obtain the band patterns shown).

individuals. For example, in Eco RI digested genomic DNA of genotype 1 (3/39 indiviuals), pAT 464 hybridized exclusively to a 4.0 kb band (Fig. 3A, lane 1); genotype 2 (27/39 individuals) included the same band as genotype 1 and an additional band of 4.5 kb (Fig. 3A, lanes 2,3,4); genotype 3 (9/39 individuals) included the same two bands as genotype 2 plus an additional band of 6.0 kb (Fig. 3A, lanes 5,6). Moreover, genotype 2 was characterized by the presence of an additional hybridizing band in digests of genomic DNA with all restriction enzymes (not shown). The size of the additional band (or allelic bands) found with a given specific restriction enzyme was invariant in all individuals of this genotype. Similar observations were made with respect to genotype 3, i.e. two additional bands were found in all restriction digests and the band sizes were invariant in individuals with this genotype. The hierarchy of hybridizing restriction fragments in the population suggests the presence of individual variation in the number of gene copies. Since all hybridizing sequences were located to a specific region of chromosome 17 (see below), this probably represents either an inconstant amplification of this chromosome region or subsequent

Table 1. Polymorphisms at the 464.2, 744.1 and 744.2 loci

Probe	Site	Enzyme	Fragment Sizes (kb)	Constant Bands	Frequency of + Site (mean ± SE)
464	464.2	Sst I	6.1 15	3.3 3.5[a]	0.29 ± .04
464	464.2	Pst I	3.6 5.2	6.8 2.6[a]	0.21 ± .05
744	744.2	Msp I	1.6, 5.7 7.2	1.4 8.6	0.24 ± .05
744	744.1	Bgl II	3.0 5.5	2.2 10.4[b]	0.22 ±.05

[a] only present in individuals with genotype 3 (i.e. containing the 464.3 locus) [b] only present in individuals with genotypes 2 and 3 (i.e. containing the 744.2 locus)

loss of the amplified region by unequal crossing over in chromosomes of some individuals. The two allele Pst I and Sst I RFLPs could be localized to the amplification unit in genotype 2 since neither allele was present in genomic DNAs of genotype 1. Other RFLPs were localized by the same method.

To assess the relative contribution of the '464.1-like' gene and the '464.2-like' gene in these complex patterns, allele specific probes were made using PCR amplification of limited segments of sequence divergence in the region 5′ of the transcription start sites of 464.1 and 464.2 (see Fig. 2 for origin of probes). Probe 1 hybridizes to both genomic clones, probe 2 hybridizes specifically to a sequence upstream of the 464.2 gene contained on genomic clone λ464.2. The patterns obtained with these probes on a Southern blot for a genotype 2 individual (Fig. 3C) clearly demonstrate the nonallelic nature of 464.1 and 464.2. A comparison of the restriction map of genomic clones λ464.1 and λ464.2 (Fig. 2), to the genomic fragment sizes in these and other digests confirms the specificity of the probes for either the 464.1 and/or the 464.2 gene. Similar to the genomic pattern seen for 464, genotype 1 individuals appear to contain a single copy of 744 per haploid genome (data not shown). An additional 744-hybridizing sequence was seen in individuals with genotypes 2 and 3 (data not shown). There was no evidence for the presence of a third copy of the 744 sequence in any individual equivalent to genotype 3 for the 464 sequence. The sizes of the bands detected by hybridizing DNA from genotype 1 individuals with a pAT 744 cDNA probe correspond to those predicted from the restriction map of genomic clone 464.1. Also, the additional 744-related gene (referred to as 744.2), is always found in those genotypes which include 464.2, and therefore, it seems likely that the two genes are linked in an amplification unit.

This suggestion is further supported by a nonrandom association of the 464.2 and 744.2 loci in population studies. We have identified restriction fragment length polymorphisms (RFLPs) for Sst I and Pst I within the 464.2 locus. Likewise, a Bgl II RFLP and an Msp I RFLP were observed for 744.1 and 744.2, respectively (see Table 1). The presence of these polymorphisms provided a method to evaluate the distance between the 464.2 and 744.2 loci. The Sst I and Pst I RFLPs at the 464.2 locus served as a useful internal control for this evaluation. As expected, highly significant linkage disequilibrium, calculated as a standardized disequilibrium constant, ϕ, was found between the Pst I and Sst I restriction site polymorphisms in the 464.2 gene (Table 2). Phage and genomic mapping studies show that the polymorphic Pst I and Sst I sites lie approximately 2 kb apart in a region adjacent and 5′ to the first exon. Somewhat surprisingly, an equally large standardized disequilibrium coefficient was found for the pairwise comparison of each of the 464.2 RFLPs with the Msp I RFLP located within the 744.2 gene indicating close genetic linkage between all three restriction sites. In contrast, there was no significant evidence for linkage disequilibrium between any of these three RFLPs and the Bgl II RFLP at the 744.1 locus. The absolute correlation between individuals containing the reduplicated genes (i.e. 464.2 and 744.2) combined with the highly significant linkage disequilibrium between the RFLPs at 464.2 and the Msp I RFLP at 744.2 strongly suggests that 464.2 and 744.2 arose from a duplication of the DNA segment that contains the linked 464.1 and 744.1 genes. The size of the reduplicated region cannot be ascertained from the failure to detect linkage disequilibrium between the 744.1 and the 744.2 genes because the absolute physical distance between two loci cannot be strictly determined from a standardized disequilibrium constant due to the fact that, among other considerations, recombination rates can vary many fold over short distances (11).

The nucleotide and protein sequence similarity between 464 and 744 and the conservation of specific amino acid residues among related genes (3) suggests that this group of genes arose

Table 2. Linkage of 464.2 and 744.2.

		Pairwise Nonrandom Association ϕ Values [a]			
		464.2	464.2	744.2	744.1
Locus	Enzyme	Sst I	Pst I	Msp I	Bgl II
464.2	Sst I	—	.628 [b]	.693 [b]	-.132 [c]
464.2	Pst I	54	—	.845 [b]	-.152 [c]
744.2	Msp I	48	48	—	-.111 [c]
744.1	Bgl II	66	62	68	—

[a] see materials and methods for the derivation of ϕ. Double heterozygotes were excluded from the analysis because the linkage phase could not be ascertained. Both the ϕ value (right upper) and the number of haplotypes examined (left lower) are shown. [b] significant at the 0.5% level. [c] not significant at the 10% level. Note: All of the polymorphisms were in Hardy-Weinberg equilibrium.

as a result of gene duplications followed by sequence divergence. To estimate the timing of the event in which the ancestral 464 and 744 genes were duplicated, generating the chromosomal region containing 464.2 and 744.2, the evolutionary distance, as measured by the number of nucleotide substitutions per site (δ), between the protein-coding regions of 464.1, 464.2 and 744.1 were compared in pairwise fashion among themselves and to the murine genes MIP-1α and MIP-1β. Assuming that the rates of nucleotide substitution are proportional to evolutionary time and that the closest genes between man and mouse (e.g. 464.1 and MIP-1α) represent 80 MY of divergence (24), the phylogenetic relationships between these genes can be described (Table 3). Similar δ values were obtained for the mouse-human pairs MIP-1α and 464.1 (δ= 0.25), MIP-1α and 464.2 (δ= 0.25) and MIP-1β and 744.1 (δ=0.22), indicating a common time of separation. The closely related human genes, 464.1 and 464.2, seem to have diverged more recently (δ= 0.029), approximately 10 MY ago (relative to mouse-human divergence 80 MY ago). Finally, the number of nucleotide substitutions between the less closely related genes, 464.1 and 744.1, (δ= 0.42), their murine analogues, MIP-1α and MIP-1β, (δ=0.44), or the cross-species pairs, MIP-1α and 744.1, (δ=0.43) and MIP-1β and 464.1 (δ=0.41), indicates that these genes duplicated in an ancestor common to mouse and man, approximately 140 MY ago.

To determine the chromosomal location of 464.1, 464.2, 744.1 and 744.2, Southern blots of Eco RI digested DNA from a panel of human-rodent somatic cell hybrids were probed with cDNA probes for pAT 464.1 and pAT 744.1. Three intensely hybridizing bands (4.0 kb, 4.5 kb, and 6.0 kb) were detected in human DNA with the 464 cDNA probe (data not shown) and two human bands (10 kb and 10.7 kb) were found with the 744

Table 3. Evolutionary distance (δ) between 464- and 744-related loci.

	464.2	MIP - 1α	MIP - 1β	744.1
464.1	0.029	0.25	0.41	0.42
464.2		0.25	ND [b]	ND
MIP - 1α			0.44	0.43
MIP - 1β				0.22

[a] see materials and methods for the derivation of δ. [b] ND indicates that the δ value was not determined for the indicated pair.

cDNA probe (data not shown). Consistent with analyses of genomic DNA band patterns in multiple unrelated individuals, a hierarchy of band patterns was also observed in the hybrid cell DNAs. The 4.5 kb and 6.0 kb bands with the 464 probe and the 10.7 kb band with the 744 probe were all absent in six independent hybrids derived from the same human donor (GM0073). The 6.0 kb band with the 464 probe was found only in 12 hybrid cell lines always in the presence of both the 4.0 kb and 4.5 kb bands, and these hybrids were derived from two human donor lines (GM0073 and VA2). Although the 464 and 744 probes clearly identified different hybridizing sequences, there was a perfect correlation for the presence or absence of both sets of bands in the panel of hybrid cell lines (Table 4). Both the 464 and 744 genes were assigned to human chromosome 17, and the genes segregated discordantly ($\geq$27%) with all other

Table 4. Segregation of 464.1 and 744.1 genes with human chromosome 17.

Human Chromosome	Gene/Chromosome +/+	+/-	-/+	-/-	% Discordancy
1	30	25	4	32	32
2	24	31	2	34	36
3	23	32	12	24	48
4	40	15	18	18	36
5	20	35	5	31	44
6	32	23	15	21	42
7	29	26	7	29	36
8	28	27	8	28	38
9	25	30	7	29	41
10	16	39	3	33	46
11	25	30	5	31	38
12	26	29	7	29	40
13	31	24	1	35	27
14	31	24	9	27	36
15	30	25	14	22	43
16	27	28	9	27	41
17	53	2	0	36	2*
18	35	20	12	24	35
19	26	29	4	32	36
20	30	25	10	26	38
21	38	17	21	15	42
22	19	36	8	28	48
X	29	26	17	19	47

The 464.1 gene was detected as a 4.0 kb hybridizing band in Eco RI digests of human-rodent somatic cell hybrid DNAs probed with a full length cDNA insert. No homologous mouse sequences were detected and faintly cross-hybridizing 4.8 and 15 kb bands were found in Chinese hamster DNA. The 744.1 gene was detected as a 10 kb band or 10 kb and 10.7 kb doublet in these same digests probed with full length 744.1 cDNA. Detection of the 464.1 and 744.1 genes, which were always both present or both absent in any hybrid cell line, is correlated with the presence or absence of each human chromosome in the group of somatic cell hybrids. Discordancy represents the presence of the genes in the absence of the chromosome (+/−) or the absence of the genes in the presence of the chromosome (−/+). The sum of these numbers divided by the total hybrids examined (×100) represents the percentage discordancy. The human-hamster hybrids consisted of 29 primary hybrids and 11 subclones (17 positive of 40 total) and the human-mouse hybrids consisted of 18 primary clones and 33 subclones (38 positive of 51 total). *The two discordancies represent detection of the hybridizing human bands in single human-hamster and human-mouse hybrids in which only a fragment of chromosome 17 was present.

human chromosomes. As expected, the bands representing the 464.2 and 744.2 genes, in addition to the less frequent reduplication of either 464.1 or 464.2 (reflected by the presence of the 6.0 kb band), were only found in the presence of the 464.1 and 744.1 sequences on chromosome 17 and did not segregate concordantly with any other human chromosome.

Chromosomal *in situ* hybridization was used to locate 744.1 more precisely to a region of chromosome 17. A total of 128 metaphases was analyzed for silver grain localization on or beside chromosomes. Of a total of 440 grains, 16 (3.6%) were clustered on q11 to q21 (Fig. 4A). This represented 61.5% of the grains present on chromosome 17 (Fig. 4B). Thirteen metaphases (10.2%) had silver grains located at 17q11-q21. No secondary peaks of hybridization were noted.

DISCUSSION

We have identified the genomic locus encoding 464.1 and 744.1, two newly described lymphokines and a nonallelic gene closely related to 464.1, which we designate 464.2. We have directly shown by genomic cloning that the 464.1 and 744.1 genes are closely linked and separated by approximately 14 kb. Previous studies have demonstrated parallel regulation of 464 and 744 following activation of T cells with various mitogens (2,3). Since the 464 and 744 genes are arranged in a head to head fashion, it seems possible that these two genes have evolved in such a way as to share some overlapping regulatory sequences. Examples have been described of regulatory sequences that affect more than one gene and act over relatively large distances (25,26). We also present evidence that the genomic segment containing the linked 464.1 and 744.1 genes is carried in a reduplicated region in some individuals; the nonallelic genes in this region have been designated 464.2 and 744.2. The 464.2 and 744.2 genes are linked as shown by the nonrandom association in unrelated individuals of RFLPs defined for these genes. The 464.2 gene is transcribed and appears to encode a translatable protein (see results). The 744.2 gene has not yet been directly characterized by cloning the genomic locus. An interesting question concerns whether individuals that genotypically carry multiple 464- and 744-related loci will show phenotypic differences relative to genotype 1 individuals (which carry only the 464.1 and 744.1 genes). Such a question can be approached with quantitative assays for the protein products encoded by these genes.

Utilizing somatic cell hybrids, we have determined that all 464-related and 744-related gene sequences map to chromosome 17. The distance between the various reduplicated loci (e.g. 464.1, 464.2 and the third hybridizing 464-related locus seen in genotype 3 individuals, referred to below as 464.3) is unknown. Because only ordered hierarchical genotypes are observed (i.e. 464.1 alone, 464.1 and 464.2, or 464.1 and 464.2 and 464.3) and no recombinant genotypes (e.g. 464.1 and 464.3) have been seen, it seems likely that the three 464-related loci are rather tightly linked and not dispersed along the length of chromosome 17. As the 744.1 and 744.2 loci are clearly linked to 464.1 and 464.2, respectively, the above rationale applies to these genes as well.

The sequence similarity between the coding segments of 464.1 and 744.1 (3), their close linkage in the human genome and their similar exon-intron structure strongly imply that these two genes have evolved from a common ancestor. Gene duplication as a mechanism for the evolution of lymphokine genes has been suggested previously by the clustering of a number of lymphokine genes (CSF2 [GM-CSF], IL3, IL4 and IL5) (27−30) on the long arm of human chromosome 5 in bands q23-q31. 464.1, 464.2 and 744.1 are related to a larger family of secreted factors (3). Absolutely conserved among the members of this family are a pair of cysteines, found either in tandem (C-C) or separated by one amino acid (C-X-C), in the first 10 to 17 residues of the processed protein. As 464.1, 464.2 and 744.1 belong to the C-C group, while a number of the C-X-C group genes (PF4, INP10 and MGSA) map to human chromosome 4q12-q21 (31−33), this family is likely to have originated from a primordial gene that dispersed to two chromosomes prior to the gene duplication and divergence which served to expand the overall number of genes. It would not be surprising to find that additional members of the C-C subgroup will be contained in the amplification unit which contains 464.2 and 744.2.

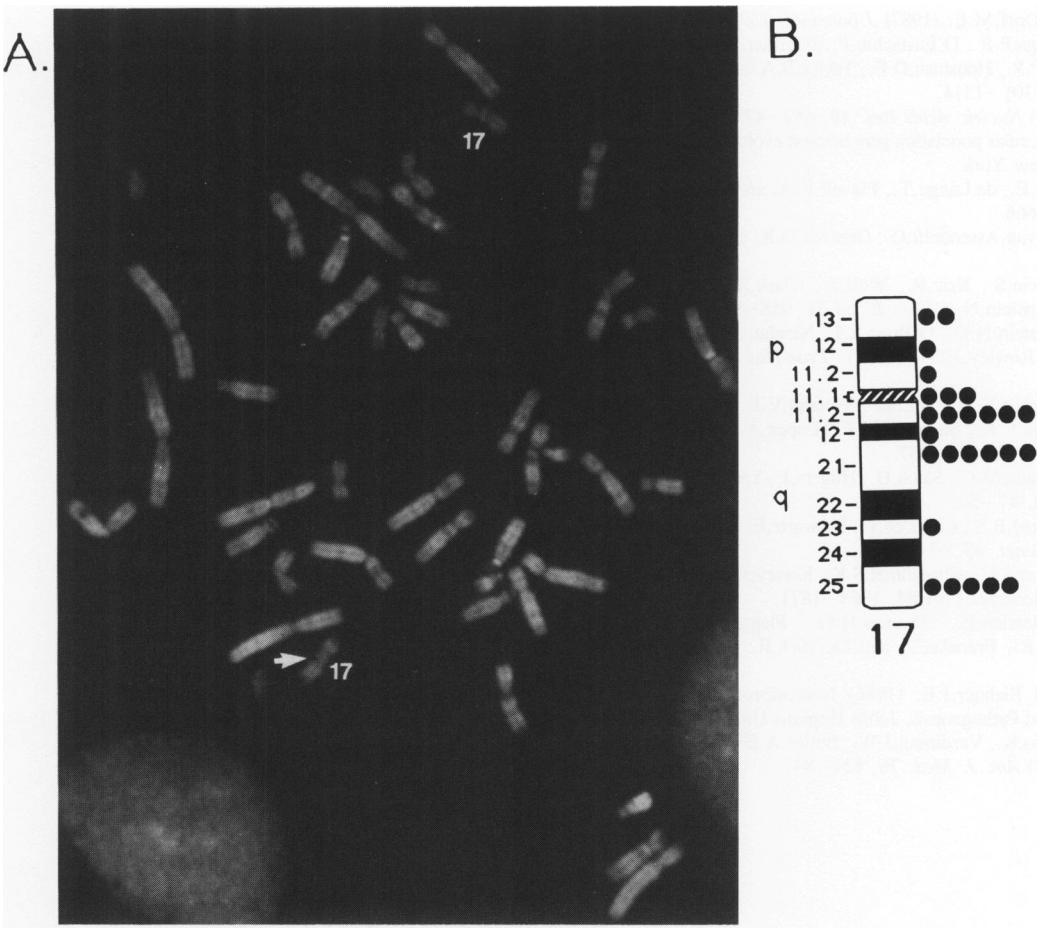

Figure 4. *Chromosomal in situ localization of 744.* Panel A shows a representative metaphase spread stained with quinacrine mustard dihydrochloride showing hybridization of the 744 probe at 17q21. Panel B shows an idiogram of human chromosome 17 with distribution of silver grains observed in 128 metaphase preparations.

Utilizing *in situ* chromosomal hybridization, we have localized the 744 locus to chromosome 17q11-q21. Genetic abnormalities in this region are strongly implicated in two human diseases, von Recklinghausen neurofibromatosis (NF1), an autosomal dominant disease (34) and acute promyelocytic leukemia (AML-M3). AML-M3 is notable in that a presumably reciprocal chromosomal translocation, t(15;17)(q22;q12-22) can be demonstrated in 90–100% of AML-M3 cases (35). The relative chromosomal locations of NF1, AML-M3 associated chromosome 17 translocation breakpoints, and the 464 and 744 loci are currently being investigated.

ACKNOWLEDGMENTS

We thank Jean Maguire and Adam Bagg for critical review of the manuscript. P. F. Z. was supported by a postdoctoral fellowship from the Deutscher Akademischer Austauschdienst. J. B. was supported as a Howard Hughes Medical Institute Research Scholar.

REFERENCES

1. Crabtree,G.R. (1989) *Science* **243**, 355–361.
2. Irving,S.G., June,C.H., Zipfel,P.F., Siebenlist,U. and Kelly,K. (1989) *Mol.Cell.Biol.* **9**, 1034–1040.
3. Zipfel,P.F., Balke,J., Irving,S.G., Kelly,K. and Siebenlist,U. (1989) *J.Immunol.* **142**, 1582–1590.
4. Wolpe,S.D. and Cerami,A. (1989) *FASEB J.* **3**, 2565–2573.
5. Wolpe,S.D., Davatelis,G., Sherry,B., Beutler,B., Hesse,D.G., Nguyen,H.T., Moldawer,L.I., Nathan,C.F., Lowry,S.F. and Cerami,A. (1988) *J.Exp.Med.* **167**, 570–581.
6. Davatelis,G., Tekamp-Olson,P., Wolpe,S.D., Hermsen,K., Luedke,C., Gallegos,C., Coit,D., Merryweather,J. and Cerami,A. (1988) *J.Exp.Med.* **167**, 1939–1944.
7. Sherry,B., Tekamp-Olson,P., Gallegos,C., Bauer,D., Davatelis,G., Wolpe,S.D., Masiarz,F., Coit,D. and Cerami,A. (1988) *J.Exp.Med.* **168**, 2251–2259.
8. Davatelis,G., Wolpe,S.D., Sherry,B., Dayer,J.-M., Chicheportiche,R. and Cerami,A. (1989) *Science* **243**, 1066–1068.
9. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc.Natl.Acad.Sci.USA* **74**, 5463–5467.
10. Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.* **132**, 6–13.
11. Chakravarti,A., Buetow,K.H., Antonarakis,S.E., Waber,P.G, Boehm,C.D. and Kazazian,H.H. (1984) *Am. J. Hum. Genet.* **36**, 1239–1258.
12. Hill,W.G. and Robertson,A. (1968) *Theor. Appl. Genetics.* **38**, 226–231.
13. Leitersdorf,E.,, Chakravarti,A. and Hobbs,H.H. (1989) *Am. J. Hum. Genet.* **44**, 409–421.
14. Jukes,T.H. and Cantor,C.R. (1969) In Munro,H.N. (ed.), Mammalian Protein Metabolism. Academic Press, New York. pp. 21–123.
15. Kimura,M. (1981) *Proc.Natl.Acad.Sci.USA* **78**, 454–458.
16. McBride,O.W., Hieter,P., Hollis,G.F., Swan,D., Otey,M.C. and Leder,P. (1982) *J.Exp.Med.* **155**, 1480–1490.
17. McBride,O.W., Battey,J., Hollis,G.F., Swan,D.C., Siebenlist,U. and Leder,P. (1982) *Nucleic Acids Res.* **10**, 8155–8170.
18. McBride,O.W., Swan,D.C., Tronick,S.R., Gol,R., Klimanis,D., Moore,D.E. and Aaronson,S.A. (1983) *Nucleic Acids Res.* **11**, 8221–8236.
19. Morton,C.C., Kirsch,I.R., Taub,R., Orkin,S.H. and Brown,J.A. (1984) *Am. J. Hum. Genet.* **36**, 576–585.
20. Rollins,B.J., Stier,P., Ernst,T. and Wong,G.G. (1989) *Mol.Cell.Biol.* **9**, 4687–4695.
21. Burd,P.R., Freeman,G.J., Wilson,S.D., Berman,M., DeKruyff,R.,

Billings,P.R. and Dorf,M.E. (1987) *J.Immunol.* **139**, 3126–3131.
22. Wilson,S.D., Billings,P.R., D'Eustachio,P., Fournier,R.E.K., Geissler,E., Lalley,P.A., Burd,P.R., Housman,D.E., Taylor,B.A. and Dorf,M.E. (1990) *J.Exp.Med.* **171**, 1301–1314.
23. Mount,S.M. (1982) *Nucleic Acids Res.* **10**, 459–472.
24. Nei,M. (1975) Molecular population genetics and evolution. North-Holland, Amsterdam and New York.
25. Kioussis,D., Vanin,E., de Lange,T., Flavell,R.A. and Grosveld,F. (1983) *Nature* **306**, 662–666.
26. Grosveld,F., Blom van Assendelft,G., Greaves,D.R. and Kollias,G. (1987) *Cell* **51**, 975–985.
27. Yang,Y.-C., Kovacic,S., Kriz,R., Wolf,S., Clark,S.C., Wellems,T.E., Nienhuis,A. and Epstein,N. (1988) *Blood* **71**, 958–961.
28. Le Beau,M.M., Epstein,N.D., O'Brien,S.J., Nienhuis,A.W., Yang,Y.-C., Clark,S.C. and Rowley,J.D. (1987) *Proc.Natl.Acad.Sci.USA* **84**, 5913–5917.
29. Sutherland,G.R., Baker,E., Callen,D.F., Hyland,V.J., Wong,G., Clark,S., Jones,S.S., Eglinton,L.K., Shannon,M.F., Lopez,A.F. and Vadas,M.A. (1988) *Hum. Genet.* **79**, 335–337.
30. Takahashi,M., Yoshida,M.C., Satoh,H., Hilgers,J., Yaoita,Y. and Honjo,T. (1989) *Genomics* **4**, 47–52.
31. Griffin,C.A., Emanuel,B.S., LaRocco,P., Schwartz,E. and Poncz,M. (1987) *Cytogenetic Cell Genet.* **45**, 67–69.
32. Luster,A.D., Jhanwar,S.C., Chaganti,R.S.K., Kersey,J.H. and Ravetch,J.V. (1987) *Proc.Natl.Acad.Sci.USA* **84**, 1868–1871.
33. Richmond,A., Balentien,E., Thomas,H.G., Flaggs,G., Barton,D.E., Spiess,J., Bordoni,R., Francke,U. and Derynck,R. (1988) *EMBO J.* **7**, 2025–2033.
34. Riccardi,V.M. and Eichner,J.E. (1986) Neurofibromatosis: Phenotype, Natural History, and Pathogenesis. Johns Hopkins Univ. Press, Baltimore.
35. Larson,R.A., Kondo,K., Vardiman,J.W., Butler,A.E., Golomb,H.M. and Rowley,J.D. (1984) *Am. J. Med.* **76**, 827–841.