Long-range G+C% mosaic domains in and around the human MHC region; characteristic genome structures and newly found genes in and around the G+C% domain boundaries, with an emphasis on the complete gene structure of GABA receptor B

Tetsushi Yamagata

Doctor of Philosophy

Department of Genetics,
School of Life Science,
The Graduate University for Advanced Studies

1997

SUMMARY

The human genome is composed of long-range G+C% (GC%) mosaic structures predicted to be related to chromosome bands. DNA replication timing, gene density, and repetitive sequence density have been connected to the chromosome band zones and the long-range GC% mosaic domains. Chromosome bands are structures that can be observed using microscopes and, when only this view is stressed, the precise location of their boundaries appears to be meaningless. However, considering the aforementioned genome features having been connected to chromosomal bands, band boundaries may be precisely assigned by placing informative landmarks (e.g. GC% transition points) on genome DNA. The human major histocompatibility complex (MHC) spans approximately 4 megabase-pair (Mbp) on the short arm of chromosome 6, and is composed of classes I (about 2 Mbp), III (1 Mbp) and II (1 Mbp) from telomere to centromere. Ikemura's group previously found that the human MHC is composed of megabase-level mosaic domains of GC%, and a boundary of the GC% mosaic domains exists in the junction area between MHC classes II and III. They predicted the GC% boundary may correspond to a band boundary. DNA replication timing during S phase is known to be correlated cytogenetically with chromosome band zones, and thus the band boundaries have been predicted to contain a switch point for DNA replication timing. One aim of my study was to determined the precise DNA replication timing for MHC classes II and III, focusing on the junction area. It was shown that the replication timing changes precisely in the GC% boundary region with a 2-hour difference of the timing during S phase, supporting the prediction that the GC% boundary is a chromosome band boundary. It was supposed that the replication fork movement terminates (pauses) or significantly slows down in the switch region, which contains dense *Alu* clusters and a long polypurine / polypyrimidine tract with the triple-helix forming potential.

The second aim of my study is to clarify the global GC% distribution in and around the entire 4 Mbp of the human MHC region. I conducted a long-range chromosome walk of about 2.1 Mbp using YACs and cosmids, and found the MHC and its surrounding region to be

composed of five long-range GC% mosaic domains disclosing three new GC% mosaic boundaries. One boundary corresponds to the junction between the MHC class II and the centromeric non-MHC region, the second does to the junction between the MHC classes III and I, and the third does to the junction between the MHC class I and the telomeric non-MHC region. It was thus shown that the GC% boundaries correspond to boundaries of functional domains of the respective genome region. In the case of the junction between the MHC class I and the telomeric non-MHC, the non-MHC region was evidently AT-rich, and the telomeric probe was located on 6p22.1 (G/Q) but close to 6p21.3 (R) by a standard fluorescence *in situ* hybridization (FISH) onto prometaphase chromosomes. Since the main body of the MHC region is present on 6p21.3, the walked region was shown to contain the boundary between 6p21.3 and 6p22.1. Characteristic structures in the GC% transition regions were searched.

The third aim of my study is to clarify genes' distribution in the walked genome region, in order to understand the chromosome bands and GC% mosaic structures on functional and evolutionary views. At first, I determine sequences of the terminal portions of the cloned fragments of individual cosmids which cover the telomeric portion of the MHC class I and the adjacent non-MHC. Because of the high density of the cosmids covering the individual regions, various portions of each of the following genes were found in a series of the consecutive cosmids. Nine genes correspond to those being previously mapped in this region, and nine genes were newly identified in this study; SMT3B-like gene (designated as SMT3X in this study), GABAB receptor gene (designated as GABR-B), mas-related gene, RASH-like gene, TRE17-like gene and MEA11-like gene, in the direction from the class I to the non-MHC, and three olfactory genes. Interestingly, sequence similarity of some of these genes with those on chromosomes 7, 11, and 17, was observed. This set of chromosomes is distinct from the three chromosomes 1, 9 and 19, on which a wide range of genes with sequence similarities with those on 6p21.3 were present. This shows that a possible boundary of the genome multiplication during evolution and/or of the genome rearrangement after the multiplication is located near the telomeric edge of the MHC class I. Elucidation of this type of boundaries and

their correlation with boundaries of GC% mosaic domains and of DNA replication timings, should give profound knowledge of evolutionary processes and mechanisms to establish the present-day human and mammalian genomes.

Among the newly identified genes in this region, I thought that GABR-B is the most interesting gene to be studied because of the following reasons, thus I determined its complete genomic sequence. This is the fourth aim of my study. The gamma-aminobutyric acid (GABA) is the most abundant and widely distributed inhibitory neurotransmitter present in the central nervous system. The receptors being activated with GABA have been classified into ionotropic (GABAA/GABAC) and metabotropic (GABAB) receptors. Two mRNA forms of the rat metabotropic gamma-aminobutyric acid receptor (GABAB receptor) have been characterized and predicted to be produced by alternative splicing. I determined a complete sequence of the human GABAB receptor gene. The pairwise alignment of this genomic sequence with that of the known rat GABAB receptor cDNAs showed the human gene spans 31 kilobases and is composed of 23 exons. The present sequence is the first example of the genomic sequence for the GABAB receptor, and thus the first study to characterize the structure for regulating transcription. The two isoforms of the GABAB receptor were found to be generated by alternative usage of promoters, rather than by alternative splicing. In both promoter regions, CpG islands and several potential transcription regulatory sites were identified. Interestingly, a cAMP response element (CRE) was present only in the promoter for the shorter isoform. Based on the fact that the GABAB receptor negatively regulates the CRE binding protein (CREB) - mediated transcriptions in the central nervous system, I proposed that the activation of GABAB receptor can differentiate the relative expression of its two isoforms through alternative usage of the promoters with and without a CRE, via the CRE- and CREB-mediated regulatory system. This may modulate synaptic transmission. An informative polymorphic marker (CA)n was found 1 kb upstream of the 5'-UTR of the GABAB receptor gene, and used to analyze 36 different HLA haplotypes. Comparison of the human GABAB receptor with related receptor proteins, such as the metabotropic glutamate receptors and the Ca^{2+}-sensing

receptors, suggest several common characteristics thought to be important for the function and/or structure of these receptors.

INTRODUCTION

1) Long-range G+C% mosaic structures and chromosome-band zones

It has become increasingly clear that the human genome, like those of warm-blooded vertebrates in general, is composed of long-range mosaic structures of G+C content (GC%), which appear to relate to chromosome bands at a high resolution level. Bernardi and his colleagues called the mosaic of large DNA regions (>300 kb on average) isochores, which are fairly homogeneous in base-composition and belong to a small number of families characterized by different GC levels (Bernardi *et al.*, 1985; Bernardi, 1989). Ikemura and his colleagues (Ikemura, 1985; Aota and Ikemura, 1986) showed a positive correlation between the G+C% (GC%) at the third codon position and the GC% of both intron and wide flanking portions, predicting long-range GC% mosaic structures of higher vertebrate genomes. Four types of chromosome bands are known for the human metaphase chromosomes produced with fluorescent dyes or differential denaturing conditions; the Giemsa or Quinacrine bands (G/Q), the ordinary Reverse bands (R), the evidently heat-stable subgroup of R bands (T), and the Centromeric bands (C) (reviewed in Comings, 1978; Therman, 1986). The G/Q and R classes are associated with a broad range of reverse functional and structural attributes. Several groups, including Ikemura's group, showed Giemsa-dark G bands to be mainly composed of AT-rich sequences, and T bands (an evidently heat-stable subgroup of R bands) of GC-rich sequences: ordinary R bands are heterogeneous and appear to be intermediate (Ikemura and Aota, 1988; Bernardi, 1989; Ikemura *et al.*, 1990; Ikemura and Wada, 1991; Bernardi, 1993; Saccone *et al.*, 1993). Furthermore, gene density, CpG island density, codon usage, chromosome condensation, DNA replication timing, repeat sequence density, and other chromosome behaviors such as recombination and mutation rates are related to chromosome bands and to long-range GC% mosaic domains (Bernardi *et al.*, 1985; Ikemura, 1985; Bird, 1987; Holmquist, 1987; Korenberg and Rykowski, 1988; Bernardi, 1989; Wolfe *et al.*, 1989;

Gardiner *et al.,* 1990; Ikemura and Wada, 1991; Bettecken *et al.,* 1992; Pilia *et al.,* 1993; Craig and Bickmore, 1994). Gene-dense R (and especially T) bands with loose chromatin structures replicate early in S phase and are rich in *Alu* repeats, while G bands with condensed chromatin structures replicate late and are rich in LINE1 repeats.

Because chromosome bands can be visualized with a microscope, precise location of their boundaries may seem meaningless. However, considering the various genome behaviors connected with chromosome bands, we can precisely locate band boundaries by placing informative landmarks on the genome DNA. Boundaries may be structurally assigned as clear GC% transition points, and functional signals may be found for punctuating and/or differentiating respective functions, e.g., a switching signal from early to late DNA replication. One aim of the present study is to solve this problem and attempt to clarify characteristics of boundaries of chromosome bands and of long-range GC% mosaic structures.

2) Mbp-level GC% mosaic domains in the human MHC region and boundaries of the G+C% mosaic domains

The human major histocompatibility complex (MHC) spans approximately 4 megabase-pair (Mbp) on the short arm of chromosome 6, and is composed of classes I (about 2 Mbp), III (1 Mbp) and II (1 Mbp) from telomere to centromere (Campbell and Trowsdale, 1993). At a standard 850-band level, the human MHC is on a T-type R band 6p21.3, but at higher resolution, a narrow G sub-band 6p21.32 is located within the MHC. For these reasons, the human MHC is a good example to study the long-range GC% mosaic domains and to examine their correlation with chromosomal band zones. Ikemura *et al.* (1988, 1990) and Fukagawa *et al.* (1995) found that the human MHC is composed of megabase-level mosaic domains of GC% possibly related to chromosome bands, and a boundary of the long-range mosaic domains is precisely located in the junction area between MHC classes II and III (Fukagawa *et al.,* 1995, 1996). All sequences from class III (spanning about 1 Mb) were evidently GC-rich.

Sequences from class II (spanning about 1 Mb), however, show significantly lower GC% levels, thus a possible boundary of the Mb-level GC% mosaic domains was assigned to the junction zone of MHC classes II and III.

DNA replication timing during S phase is known to be related cytogenetically with chromosome band zones, and thus the band boundaries have been predicted to contain a switching point for DNA replication timing. If the Mb-level GC% mosaic boundary noted above correspond to a boundary of chromosome band zones, the DNA replication timing may switch in the GC% transition region. In order to examine whether the junction of classes II and III corresponds to a switching point of the DNA replication timings during S phase, Dr. Tenzen and I determined the replication timings for the respective region at the nucleotide sequence level (Tenzen *et al.*, 1997).

3) Genes identified around the junction area of MHC classes II and III

The distribution of genes in distinct band zones is strikingly non-uniform (Bernardi *et al.*, 1985). Gene concentration in T bands with high GC% levels is predicted to be about twenty times higher than in G/Q bands with low GC% levels, and several times higher than in ordinary R bands with intermediate GC% levels (Mouchiroud *et al.*, 1991). This general tendency was confirmed during studies to characterize the GC% domain border in the junction between MHC classes II and III: in this study, Ikemura and his colleagues found four new genes in the GC-rich class III side (Matsumoto *et al.,* 1992a, 1992b; Sugaya *et al.,* 1994); TNX, an extracellular matrix protein tenascin-like gene; AGER, a receptor gene for advanced glycosylation end products of proteins; PBX2, a possible proto-oncogene with a homeobox sequence previously called HOX12 by Sugaya *et al.* (1994); and NOTCH4. Furthermore, from these results, Sugaya *et al.* (1994) noted linkage similarity between two sets of genes located on 6p21.3 and on 9q32-34: the human gene closely related to NOTCH4 is NOTCH1 (TAN1) which is located

on 9q34.3; the gene related to PBX2 is PBX3 mapped on 9q33-34; and the gene related to TNX is HXB (the tenascin C gene) on 9q32-q34 (see Table 1). This linkage was predicted to be formed by segmental genome duplication during evolution. If this linkage similarity between the noted sets of genes really reflects part of the paralogous chromosomal segments caused by segmental genome duplication during evolution, then there should exist other pairs of genes with analogous characteristics in the respective regions. Actually, 10 more genes on 6p21.3 have been found to have counterparts mostly mapped on 9q33-q34 (Kasahara *et al.*, 1996; see also Table 1). Detailed analysis on the mouse gene organization has also revealed a similar genome duplication between mouse chromosome 17, harboring the murine MHC, and chromosome 2 (Kasahara *et al.*, 1996). More recently, the similarity among human chromosomes 1, 6, 9 and 19 has been noted by Katsanis *et al.* (1996). For example, genes homologous for NOTCH4, PBX2, and TNX aforementioned were mapped not only on chromosome 9 but also on chromosomes 1 and 19: NOTCH2 on 1p13-11, NOTCH3 on 19p13.2-13.1, PBX1 on 1q23, and TNR (the tenascin R gene) on 1q25-31. Based on the available data in the current Genome Database, Sugaya *et al.* (1997) summarizes the sets of homologous genes mapped on these human chromosomes; Table 1 is from the paper of Sugaya *et al.* (1997). These observations should have implications not only for understanding the origin of the present-day MHC but also for understanding the processes to be involved in the vertebrate genome evolution, especially when another example for such genome multiplication around MHC is found. The newly identified genes near the telomeric end of the class I region in the present study disclosed the linkage similarity with the genes belonging another set of chromosomes. Since these chromosomes are totally different from the set noted above, there appears to exist a boundary of the genome multiplication and/or of the rearrangement after the multiplication.

MATERIALS AND METHODS

1) Construction of cosmid libraries and isolation of cosmid clones.

Four cosmid libraries were constructed with the Supercos-1 (Stratagene) cosmid vector from total DNA isolated from the yeast strains (AB1380) harboring the four YACs which cover the human MHC class I region; one strain contains CEPH mega-YAC 745D12 with a 570-kb insert, the second contains 960H11 with a 1.6-Mb insert, the third contains YAC 788E2 with a 960-kb insert, and the fourth contains YAC 800G3 with a 960-kb insert. These four YACs have been shown not to be chimeric (Okumura, K., *et al.*, unpublished data and Burt et al., 1996). Total recombinant yeast DNA was partially digested with *Sau*3AI, ligated to the Supercos-1 vector, and packaged with Gigapack II gold packaging extract (Stratagene). Replicas were prepared by spreading colonies from each cosmid library on Hybond N^+ filters and hybridized with radio-labeled human genome DNA as described below. The isolated cosmids containing human DNA inserts were transferred to 96-well microtiter plates, cultured, and spotted on Hybond N^+ filters. The filters arraying 1008 clones were denatured, neutralized, baked at 80°C for 2 h, and hybridized at 65°C with ^{32}P-labeled DNA probes of interest in 5 × SSPE, 5 × Denhardt's solution, 100 μg/ml freshly denatured salmon sperm DNA, 20 μg/ml freshly denatured human placental DNA, and 0.1% SDS. After hybridization, the filters were washed twice at 65°C with 2 × SSC and finally with 0.1 × SSPE containing 0.1% SDS, and autoradiographed at -80°C with intensifying screens.

2) Chromosome walking; Cosmid Isolation and contig construction.

Cosmid contigs were assembled by colony hybridization in arrayed-libraries with a representation of three to seven YAC equivalents, using a restriction-enzyme digested fragment of an already obtained clone as its probe, which was labeled with [α-^{32}P] dCTP by the random-priming method. The preliminary contigs were constructed using two cosmid clones (pM67 and pM78) and PCR-product probes corresponding to the loci previously mapped in the

MHC class I region: S, HSR-I, HLA-L, -G, -F, MOG, RFP, D6S265 and D6S131. Hybridization with the labeled probe DNA was performed in hybridization buffer containing 5 × SSPE (1 × SSPE is 0.18 M NaCl, 1 mM EDTA, and 10 mM NaH_2PO_4 pH 7.7), 5 × Denhardt's solution (1 × Denhardt's solution contains 0.02% Ficoll, 0.02% polyvinylpyrrolidone, and 0.02% bovine serum albumin), 100 μg/ml freshly denatured salmon sperm DNA, 20 μg/ml freshly denatured human placental DNA, and 0.5% sodium dodecyl sulfate (SDS) at 65°C for 15 to 18 h after prehybridization in the same buffer solution without the probe DNA as described by Sealey *et al.* (1985). Continual cosmid contigs, which were located between BAT1 and RFP in the human chromosome 6p21.3-22.1 were constructed, and the $GABA_B$ receptor gene was located in the MHC class I region.

3) Restriction mapping and Southern blot analysis of cosmids.

Cosmid clones were cultured in 2.0-ml LB medium containing 20 μg/ml of ampicillin for 17 hours at 37°C. Cloned DNAs were purified with a plasmid isolation system (Kurabo Model PI-100∑). A 0.2-0.3 μg of each cosmid was digested with *Eco*RI and then resolved in a 0.7% agarose gel. Southern transfers were performed onto Hybond-N^+ filters as described by Sambrook *et al.* (1989). Hybridization was performed as described above.

4) Sequencing of both ends of cosmid-cloned fragments.

A 0.1-0.2 μg of each cosmid was digested with *Bam*HI or *Xho*I. The fragment was self-ligated with T4 DNA Ligation Kit Ver. 2 (Takara Shuzo Co., Ltd., Kyoto, Japan). A fragment with the vector sequence was amplified by PCR with T7 primer (5'-CCGCATAATACGACTCACTATAGG-3') or T3 primer (5'-GCCGCAATTAAC CCTCACT AAAG-3'), designed from the cloning site of the vector. The PCR conditions included the following: 2.5 U Ex Taq polymerase (Takara Shuzo Co., Ltd., Kyoto, Japan), 0.4 uM primers, 1 × Ex Taq Buffer (final concentrations); denature at 94°C for 1 min, 16 cycles at 94°C for 15 sec, 68°C for 10 min and then autoextending 15 sec per cycles for 16 cycles, followed by 72°C

for 10 min. The DNA was sequenced by the Taq cycle sequencing kit, using standard protocols of the ABI PRISM™ Dye Terminator Cycle Sequencing kit for the ABI377 automated sequencer (Applied Biosystems, Foster City, CA).

5) Genome DNA Sequencing and rapid amplification of cDNA ends (RACE).

Fragments from cosmids, produced by digestion with restriction enzymes (*Bam*HI, *Bgl*II, *Eco*RI, *Eco*RV, *Hind*III, *Hinc*II, *Pst*I, *Sac*I, *Sma*I, *Sph*I, *Xba*I, and *Xho*I), were cloned into pUC118 and then transformed into *E. coli* JM109 or STBL2 (GIBCO BRL). The 5.0 kb-*Eco*RI and 6.2 kb-*Bam*HI fragments containing the second promoter region, which is located between exon 4 and 5, were difficult to subclone using JM109, and the recipient cell for these transformations was changed to STBL2. These subclones were sequenced using standard protocols of the ABI PRISM™ Dye and/or BigDye Terminator Cycle Sequencing FS Ready Reaction Kit for the ABI377 automated sequencer. The following oligonucleotides were used for the sequencing primers: -21M13, 5' GTTGTAAAACGACGGCCAGT 3'; M13Rev, ACAGCTATGACCATGATTAC 3'; 3'GABAPromoter, 5' GGCTGGAGCCTGGATTCGA GGGGA G 3'; 5'Bg10SM1new, 5' GGAGGAGAGAAAGCCTGTCCCCAC 3'; 3'Bg10SM1new, 5' CCAACAAAATCAG GGATGGAGGCGC 3'; 5'Exon5, 5' CAGCCCC CGCTCATGGGAAACAG 3'.

The 3' RACE was carried out according to the manufacture's recommendation (GIBCO BRL) using 500 ng of poly A$^+$ human brain mRNA (Clontech). Exon 21 primer, 5'-CCGTGAACTGGAAAAGATCATTG-3', was used as the gene specific primer. Amplified products were cloned to pT7Blue T-vector (Novagen) and sequenced as described above.

6) Sequence data analysis and data-submission.

Nucleotide and amino acid sequences were analyzed using the computer programs GENETYX-MAC™ and DNASIS™. Repetitive sequences were classified by RepeatMasker2 program (Smit, A. F. A. & Green, P., http://ftp.genome.washinton.edu/cgi-bin/RepeatMasker). A

database search in DDBJ, GenBank, EMBL, PIR, and SWISS-PROT was performed by FASTA and BLAST programs (Pearson and Lipman, 1988; Altschul *et al.*, 1990), and multiple alignment of amino acid sequences was determined by the CLUSTAL W program (Thompson *et al.*, 1994). A prediction of complete gene structures and of promoters was made with the GENSCAN (Burge and Karlin, 1997) and NNPP program (Reese *et. al.*, http://www-hgc.lbl.gov/progects/promoter.html), respectively. The nucleotide sequences for the $GABA_B$ receptor gene (GABR-B) and for SMT3X were deposited in DDBJ/GenBank/EMBL under the accession numbers AB009982 and AB009983, respectively.

7) Micro-satellite polymorphism analysis.

Genomic DNAs were extracted from 36 Japanese HLA homozygous B-lymphoblastoid cell lines. PCR primers for determining the number of the (CA) repeats were as follows: 5'CA-R, 5'-CCAACAAGGAGCCCACTGTTTCCTC-3'; and 3'CA-R, 5'-CTCTAAAGACAGCAGGT ACAGACC-3'. Amplification was carried out in 50 μl containing 40 ng of genomic DNA and 10 pmol of each primer. After initial denaturation at 95°C for 5 min, amplification was for 35 cycles with denaturation at 95°C for 20 s, annealing at 58°C for 15 s, with an extension at 72°C for 30 s, followed by a final extension at 72°C for10 min. The PCR products were purified with SUPREC™-02 (Takara Shuzo Co., Ltd., Kyoto, Japan) and used as the template for sequencing with each PCR primer as the sequencing primer.

8) GC% measurement of DNA fragments.

Insert DNAs (30-40 kb) derived from cosmid clones were separated from RNA and the vector DNA by high-performance liquid chromatography (HPLC); cosmid DNAs were digested by *Not*I at the Supercos-1 linker, put on a TSKgel DEAE-NPR HPLC column (0.46 × 3.5 cm; Tosoh Co., Tokyo) and eluted with a linear gradient of NaCl (from 0.5 to 1 M) in 0.02 M Tris-HCl (pH 9.0). GC% of purified inserts was measured with the DNA-GC kit (Yamasa Shoyu Co., Chiba, Japan) according to the manufacturer's protocol; 20 mg of DNA (EDTA free) dissolved in 20 μl of distilled water was heated at 100°C for 5 min followed by rapid cooling in

an ice bath, mixed with 20 ml of nuclease P1 solution (2 units/ml of 40 mM sodium acetate buffer containing 0.2 mM $ZnCl_2$, pH 5.3), and incubated at 50°C for 1 hour. The P1 hydrolysates and a standard mononucleotides mixture, supplied by the manufacturer, were separately chromatographed on a YMC reversed-phase HPLC column (ODS-AQ-312, 0.6 x 15 cm; YMC Co., Kyoto) in 10 mM H_3PO_4 - 10 mM KH_2PO_4 (pH 3.5) at 26°C, and GC% was calculated according to the manufacturer's protocol.

9) Chromosome *in situ* hybridization.

Fluorescence *in situ* hybridization (FISH) was used to assign the chromosomal location of the cosmid TY6G11. Chromosome spreads were obtained from phytohemagglutinin-stimulated blood lymphocytes of a healthy male donor. The cosmid was labeled with biotin-16-dUTP (Boehringer Mannheim) by nick translation. *In situ* hybridization was performed according to standard procedures in the presence of COT-1 DNA as a competitor. The hybridized probe was detected with FITC-conjugated avidin (Boehringer Mannheim) without further signal amplification. Chromosomes were counterstained with 0.2 μg/ml propidium iodide for R-banding. Fluorescence signals were imaged using a Zeiss Axioskop epifluorescence microscope equipped with a cooled Charge Coupled Device (CCD) camera (Photometrics, PXL 1400). Image acquisition was performed on a Macintosh computer with the software program IPLab™ (Signal Analytics Co.). The images were then pseudocolored and merged using Adobe Photoshop™ 3.0J (Adobe Systems Inc.). Hoechst, FITC, and propidium iodide images were shown in blue, green, and red, respectively.

10) Cell culture, labeling, and isolation of newly replicated DNA.

Human myeloid leukemia HL60 cells were grown as suspension cultures in RPMI medium 1640 supplemented with 10% fetal calf serum (GIBCO BRL). The cell cultures were labeled with 0.1 μM [^{14}C]thymidine (2.11 GBq/mmol; Amersham) for 3 days and synchronized at the G1/S boundary by two successive treatments with aphidicolin (1st, 1μg/ml; 2nd, 5μg/ml), an inhibitor of DNA polymerase α, as detailed by Tribioli *et al.* (1987) and Pedrali-Noy *et al.*

(1981). After washing with a fresh medium to release the block, cells were labeled at intervals of one hour with 75 μM bromodeoxyuridine (BrdUrd), an artificial analogue of thymidine, and 0.04 μM [^{3}H]deoxycytidine (925 GBq/mmol; Amersham) from the onset of S phase until the sixth hour. Cells were killed by 0.02 % NaN_3, collected by centrifugation, and kept at -70°C. Newly replicated DNA was purified according to Yoon *et al.* (1995) and Contreas *et al.* (1992). Nuclei isolated from the labeled cells were loaded directly onto 5% to 30% alkaline sucrose density gradients for size fractionation, and the pooled fractions containing short DNAs with high ^{3}H but low ^{14}C counts, which were mainly less than 5 kb in length, were subjected to CsCl density gradient centrifugation in a Beckman SW50.1 rotor (42,000 rpm for 48 hours at 21°C), according to Yoon *et al.* (1995). The fractions with high ^{3}H but low ^{14}C counts were again pooled and dialyzed against TBSE (10 mM Tris-HCl pH7.5, 150 mM NaCl, 0.1 mM EDTA). Nascent DNA was further purified on an immunoaffinity chromatography column with anti-BrdUrd antibodies (Contreas *et al.*, 1992). The elute was treated with proteinase K, phenol, and chloroform followed by ethanol precipitation. All steps were performed in the dark.

11) Quantification of nascent DNA by competitive PCR.

Competitor DNAs for individual genome sites were prepared using the method described by Diviacco *et al.* (1992), ligated with pT7Blue T-vector (Novagen), and transformed into *E.coli.* JM109. Plasmids obtained were linearized with *Eco*RI, and were quantified by spectrophotometric measurements. A constant ^{3}H-dpm of nascent DNA from individual time periods was co-amplified with a constant amount of competitor DNA in a single tube, by PCR of 32 cycles for 15 seconds at 94°C, 15 seconds at 56-60°C, and 20 seconds at 72°C in a 9600 Thermal Cycler (Perkin Elmer). Several different PCR conditions (annealing temperature and ratio between competitor and nascent DNA) were tested. Each amplified product from the nascent DNA was 190-200 nt in length, and the competitor DNA had an additional 17-nt insertion in the middle of the amplified region. PCR products were separated on 8% polyacrylamide gels and stained with fluorescent dye CYBR GREEN I (Molecular Probes

Inc.). Bands were visualized and quantified with a FluorImager SI (Molecular Dynamics). The amount of nascent DNA was then calculated by comparison with the band derived from the respective competitor.

RESULTS

1) Characterization of a boundary of long-range GC% mosaic domains identified in the junction of MHC classes II and III.

The chromosomal region containing the major histocompatibility complex (MHC) is the best described region of the mammalian genome in both its structure and genetics. One aim of my study is to clarify long-range GC% mosaic domains in the MHC, to assign both the GC% boundaries and the replication switch region, and to find characteristic structures present in the boundary region; i.e., to characterize the boundary of chromosomal band zones at a nucleotide sequence level. In the junction of MHC classes II and III, Fukagawa *et al.*, (1995) reported a sharp GC% transition and proposed the transition may correspond to a chromosome band boundary. In the first part of my thesis, I described the study which showed that the GC% mosaic boundary corresponds to the replication timing switch region. Then I described global GC% distribution in and around the entire 4 Mbp-MHC region by conducting a long-range chromosome walk, a total of about 2.1 Mbp. The MHC and its surrounding region was found to be composed of five long-range GC% mosaic domains disclosing three new GC% mosaic boundaries.

Another aim of my doctor course study is to determine the gene distribution in the walked region to correlate the GC% mosaic domains with the genome domains defined by functional aspects. This may enable us to understand the evolutionary processes and mechanisms responsible for formation of the present mammalian genome. As a model system, I will focus on the MHC and its surrounding regions. In the walked region, I found six new and three known genes which showed sequence similarity to genes on chromosomes 7, 11, and 17. Since these chromosomes were different from those previously described for the linkage similarity with the genes on 6p21.3, we have new insight to understand the evolutionary processes involved in the genome organization of this region and thus of the MHC.

1. Replication timing for the GC content transition area

Mammalian DNA replication timing is roughly divided into early S (S_E) and late S (S_L) phases; G band zones mainly replicate in S_L, whereas R band zones, including T bands, replicate in S_E (Holmquist *et al.*, 1982). Band boundaries have been predicted to correspond to switching regions of DNA replication timings (Drouin *et al.*, 1994), and therefore have been presumed to be rather precisely assignable by locating the switching point of replication timing from early to late, where the regulatory signals for DNA replication would exist. In addition, band boundaries should be suitable for studying mechanisms of termination, pausing, and/or slowing down of DNA replication fork movement. As noted in the INTRODUCTION, Ikemura and his colleagues have reported that the MHC is composed of Mb-level GC% mosaic domains (Ikemura and Aota, 1988; Ikemura *et al.*, 1990). By conducting bi-directional chromosome walking of a 450-kb region (Fukagawa *et al.*, 1995; Matsumoto *et al.*, 1992b; Sugaya *et al.*, 1994) spanning HLA-DRA (the AT-rich class II side) and CYP21 (the GC-rich class III side), Fukagawa *et al.* (Fukagawa *et al.*, 1995, 1996) identified a boundary of the Mb-level GC% mosaic domains disclosing a sharp GC% transition, and predicted this transition region would be an example of a chromosome band boundary. Dr. Tenzen and I studied the replication timing in the MHC region focusing on the junction of class II and III (Tenzen *et al.*, 1997). In the measurement of the replication timing using competitive PCR, DNA sequences of the examined loci were required, and I determined the sequence.

Figure 1 shows the genome structure in the 450-kb region where the bi-directional chromosome walking was conducted (Fukagawa *et al.*, 1995). To examine replication timing for specific regions during S phase, we used the method based on accurate quantification of the newly replicated DNA of the regions of interest by competitive PCR. This method is suitable for examining absolute replication timing for genome sites at the nucleotide-sequence level. Human myeloid leukemia HL60 cells that had incorporated [^{14}C]thymidine in advance were synchronized at the G1/S boundary by two successive treatments with aphidicolin. S phase was started by removing aphidicolin and cells were labeled with both [^{3}H]deoxycytidine and BrdUrd at 1-hour intervals with the purpose of examining replication timing for Mb-level

mosaic domains, each of which is most likely composed of multiple replicons. Incorporation of [^{3}H]deoxycytidine was found to continue at high levels for six hours after the onset of S phase; then it dropped to about one fourth during the next two hours, showing good cell synchronization. In the present study a labeling time of one hour was chosen to investigate the replication timing for the Mb-level GC% mosaic domains, each of which presumably consists of multiple replicons coordinately replicating in a certain period of S phase. In the present study, samples from the first six hours were used. The ^{3}H- and BrdUrd-labeled short DNA (mainly less than 5 kb) for each time interval was separated from the long and light ^{14}C-labeled genomic DNA as described in MATERIALS AND METHODS.

To precisely quantify a small amount of the nascent DNA corresponding to the genome locus of interest, competitive PCR (Diviacco *et al.*, 1992; Gilliland *et al.*, 1990; Siebert *et al.*, 1992) was employed. Using each of the fifteen sets of PCR primers (Table 2), which gave a single band for the human genomic DNA, nascent DNA from six time intervals was co-amplified with the corresponding competitor DNA with a 17-nt insertion in the middle of the amplified region, followed by electrophoresis on polyacrylamide gels (Fig. 2). For all four sites located in the 180-kb AT-rich region of the class II side (DRA, CP, PL, LINE CLUSTER), the fourth time interval showed the strongest band for the BrdUrd-substituted nascent DNA, whereas the intensities for competitor bands were practically constant for all the time intervals. Therefore these loci mainly replicate during the fourth hour of S phase. The amounts of nascent DNA corresponding to individual sites were quantified by comparison of nascent DNA bands with competitor bands, performing three independent PCR (Fig. 3). In two sites located in the 150-kb class III region with higher GC% (TNX, AGER), the second time interval was amplified most effectively (Figs. 2 and 3), showing that these sites replicate mainly during the second hour of S phase and thus about two hours earlier than the AT-rich class II side. Another class III gene TNFA, which is also GC-rich and about 500 kb away from TNX (Fig. 1), was found to replicate during the first interval of S phase, consistent with the relationship that a GC-rich region replicates early.

As the major transition of replication timing was localized within the 100-kb region between LINE CLUSTER and AGER (Fig. 4), eight PCR primer sets were designed to cover the intervening region more densely (PN112, PCD, NA, SR, NCT, CTG REPEATS, INT3A, INT3B). The patterns were intermediate between the two observed for classes II and III (Figs. 2, 3). Figure 4 shows that the transition point for replication timing can be precisely assigned and is highly correlated with the GC% distribution. The mid-point of the replication transition is located about 220 kb telomeric from DRA, i.e. between the *Alu* rich region and NOTCH4 (a human counterpart of mouse Int3 gene). The mid-point precisely coincides with the major GC% transition region. The replication timing appears to differ by one hour between PCD and INT3A, which are 16 kb apart from each other. The finding that replication timing changes precisely in the GC% transition region supports our previous prediction that the boundary of the long-range GC% mosaic domains in the MHC corresponds to the chromosome band boundary (Fukagawa *et al.*, 1995).

Using each of the thirteen sets of PCR primers listed in Table 2, nascent DNA from six time intervals was co-amplified with the corresponding competitor DNA followed by electrophoresis on polyacrylamide gels (Fig. 2). For all four sites (DRA, CP, PL and LINE CLUSTER) located in the 180-kb AT-rich region of the class II side, the fourth time interval showed the strongest band for the BrdUrd-substituted nascent DNA, indicating that these loci mainly repl3HcaHlified most effectively (Figs. 2 and 3),
indicating that these sites replicate mainly during the second hour in S phase and thus about two hours earlier than the AT-rich class II side. Another class III gene TNFA, which is GC-rich and

about 600 kb telomeric from TNX (Fig. 1), was found to replicate during the first interval of S phase, again showing a GC-rich region belongs to an earlier replication zone.

In the 25-kb region which contains the sharp GC% transition region and partly overlaps the dense Alu clustering region (Fig. 1), five sites (PN112, PCD, NA, SR and NCT) were examined. As the replication switching point was localized within this 25-kb region in the course of the present study, I sequenced the region intensively in order to design the PCR primer sites for the region more closely than for other regions. For the two sites, PN112 and PCD, relatively equivalent levels of PCR products were observed in the third and fourth time intervals (Figs. 2 and 3), showing both sites replicate about 30 minutes earlier than the four AT-rich class II sites aforementioned (from DRA to LINE CLUSTER). Three sites, NA, SR and NCT, replicate during the third hour (Figs. 2 and 3), and thus an approximately one-hour lag was present between NCT and the aforementioned CTG REPEATS belonging to the earlier replication zone, which are only 6 kb apart from each other. Figure 4 shows that the switching point for the replication can be precisely assigned and the replication timing is highly correlated with the long-range GC% distribution. The finding that replication timing switches precisely in the GC% transition region supports our previous prediction that the boundary of the long-range GC% mosaic domains in the MHC is an example of chromosome band boundaries (Fukagawa *et al.*, 1995). We propose the methods used in this and the previous studies as a general strategy for defining band boundaries at the nucleotide level and placing valuable landmarks on the human genome, which will help us to comprehensively understand molecular genetic and cytogenetic observations of the human genome.

There is a two-hour gap of replication timing between LINE CLUSTER (the class II-side) and CTG REPEAT (the class III-side), which are about 30 kb apart (Fig. 4). Interestingly, a one-hour difference in replication timing was observed between two sites (NCT and CTG REPEAT) which are only 6 kb apart (Fig. 4); the replication from CTG REPEAT to NCT is thought to terminate or to pause and it was designated as the major switching region. A total of a two-hour gap in replication timing between classes II and III, however, may not solely be due

to this single termination (or pausing) site since replication timing appears to slightly change from LINE CLUSTER to NA. The replication fork movement may pause at multiple sites or traverse very slowly there, e.g., one tenth of the normal speed of 30-50 nt per second (Huberman and Riggs, 1968; Falaschi et al., 1993). The relatively broad peak observed for the sites from PN112 to CTG REPEATS (Fig. 3) may support this model.

2. Characteristic structures found in and around the G+C% domain boundaries identified in the junction of MHC classes II and III

It should be noted that in the case of GC% transitions, very precise assignment such as within a few kb may be difficult because local GC% fluctuations exist due to local genome structures such as CpG islands with high GC% levels. Fukagawa *et al.*, (1995) previously defined the border of Mb-level GC% domains in the MHC at the cosmid level by size (e.g. several ten kb) and thus designated it the GC% transition region rather than the transition point. The present study shows that the switching point for DNA replication timing can be more accurately assigned and the major transition zone has already been localized within a several kb region. While molecular mechanisms of replication arrest have been reported in *Escherichia coli* (Hidaka *et al.*, 1988) and *Saccharomyces cerevisiae* (Kobayashi *et al.*, 1992), little is known about the way in which the replication of higher eukaryotes terminates or pauses. In humans, termination of the replication fork was identified in rRNA genes (Little *et al.*, 1993), and the termination/pausing region of the replication fork was identified, which encompasses a dense *Alu* cluster and polypurine/polypyrimidine tracts (Little *et al.*, 1993).

The switching region identified in the present study also contains characteristic sequences, which we assume to be candidates for causing the termination (or pausing) of replication fork movement. One of them is a 210-bp polypurine/polypyrimidine tract (pur/pyr tract). It is worth noting that the replication fork movement is known to be arrested at pur/pyr tracts *in vitro* (Baran *et al.*, 1991, Brinton *et al.*, 1991) although the function *in vivo* has been scarcely examined. Figure 5A shows the 210-bp pur/pyr tract found in the human MHC, which is

present at the exact mid point of the replication-timing switch zone with 2-hours difference during S phase, and of the GC% transition (Fig. 4). Its purine strand can be mostly expressed as palindromic stretches composed of the two tetranucleotides, (GGAA) and (AGAA). Under the proper superhelical strain, palindromic pur/pyr sequences (termed H-palindromes) undergo a distinct structural transition involving an intrastrand disproportionation, in which one half of the mirror-repeat sequence dissociates into single strands and one of the two single strands folds back on the other half of the mirror-repeat duplex to form an intramolecular triplex (known as H-DNA). Figure 5B shows an example of such H-DNA triplexes formed under a neutral pH condition for the 210-bp pur/pyr tract when only the perfectly consecutive Hoogsteen-pairing was assumed. H-DNA was originally found as the triplex being formed at the mirror repeat sequences in supercoiled plasmids (Wells *et al.*, 1988), and the present 210-bp pur/pyr tract was shown to form triplex efficiently in a supercoiled plasmid *in vitro* (an unpublished result of T. Tenzen and T. Ikemura).

Several other characteristic structures was also found in and around the switching region. Figure 6 shows the sequence for the 5.7-kb *Eco*RI fragment. This sequence is highly AT-rich (70%) and rich in A_{3-5} or T_{3-5} tracts being spaced about 10 nt between the tracts (Fig. 6). These characteristics are often seen in the canonical SARs isolated to date. The A+G% distribution in Fig. 6 also shows that the 5.7-kb region is evidently biased in the purine/pyrimidine ratio. One reason for this bias is the dense clustering of *Alu* element whose purine/pyrimidine ratio is known to be biased. Besides *Alu* elements, there also exists clustered polypurine/polypyrimidine tracts: only the complete forms of the polypurine/polypyrimidine tracts are listed in Fig. 6, and much longer but slightly imperfect polypurine/polypyrimidine tracts are also present in this region. Figure 8 lists other characteristic sequences in and around the switching region; for example, di-, tri- and tetranucleotide repeats (e.g. GAA, CAG, ATT, and GGAA repeats) and new types of medium reiteration frequency sequences (MERs; Jurka *et al.*, 1993) were found around the switching region (Fig. 7). There also exists *Alu* and LINE1 clusters as previously reported (Fukagawa *et al.*, 1995 and Sugaya *et al.*, 1994).

Another characteristic structure present in the vicinity of the GC% transition region is a pseudoautosomal boundary-like sequence (PABL). Fukagawa *et al.* (1995) defined PABL as an approximately 650-nt sequence, highly homologous (about 80% nucleotide identity) with human pseudoautosomal boundary (PAB), which is the interface between sex-specific and pseudoautosomal regions (PAR) in the sex chromosomes (Ellis *et al.*, 1989; Ellis *et al.*, 1989; Ellis *et al.*, 1990). The sex chromosomes are divided into two functionally distinct regions, sex specific sequences and pseudoautosomal regions (PARs), and the PAB of the Y chromosome (PABY) has been reported to position itself near a boundary of long-range GC% mosaic domains (Whitfield *et al.*, 1995). It should be noted that when DDBJ/EMBL/GenBank was searched using the aforementioned 210-bp sequence in the MHC as a query, about ten human genomic sequences with high homology were found. Most of them were in the long sequences determined by large-scale sequencing and were poorly characterized, but one was found in the vicinity of a well characterized locus "pseudoautosomal boundary (PAB)" on the short arm of the human sex chromosomes. The pur/pyr tract identified near PAB is about 800 bp long and is composed of several distinct domains; e.g., a domain mainly consisting of (GGAA/TTCC) and (GGGA/TCCC) systematically arranged, another domain composed of (GGAA/TTCC), (GAAA/TTTC) and (GGGA/TCCC) again systematically arranged, as well as more complex domains of about 300 bp containing a 5% level of pyrimidine nucleotides. It should be noted that these two GC% boundaries have analogous genome characteristics possibly related with chromosomal band boundaries at a high resolution level.

2) Mbp-level GC% mosaic domains in and around the MHC region

The chromosomal region containing MHC is the best described region of the mammalian genome in both its structure and genetics. Our goal of Ikemura's group is to increase the knowledge of this region and use it as a paradigm for the structure and evolution of the mammalian genomes.

1. Chromosome walking from MHC class II to non-MHC region.

Human MHC class II region, and thus the class II genes, has been extensively cloned and analyzed by many groups. In order to clarify the GC% distribution in and around the class II region, 17 cosmid clones obtained from Tokai Univ., and 10 clones (AK sires) which I isolated from the human genomic library constructed by Tokai Univ., were analyzed. Some of the clones were shown to be located near and at the COL11A2 locus which is known to be the immediately centromeric neighbor to the class II region (Fig. 8). Cosmid AK1G5 which was predicted to contain the junction between the MHC class II and the centromeric non-MHC region, was subcloned and sequenced. The GC% distribution was also obtained from the 198-kb class II sequence which was reported by Beck *et al.* (1996). Figures 9A and B show that a boundary of long-range GC% domains is present in the junction between the MHC class II and the centromeric non-MHC region. Again, the GC% boundary corresponds to a boundary of functional domains.

2. Chromosome walking from MHC class I to non-MHC region, and fine mapping of the walked region

While the proximal half of the MHC region corresponding to the classes II and III has been extensively studied, the distal half of the MHC which corresponds to the class I region has been less characterized. This is partly because many of the MHC class I family members are pseudogenes or gene fragments. Among the six loci expressed in human, HLA-A, -B, and -C heavy chains are expressed in a wide range of tissues, and are associated with ß2-microglobulin and present peptide antigens to T cells. The heavy-chain products of the additional loci, HLA-E, -F, and -G, also associate with ß2-microglobulin but their function are less clear. The multigenic family of MHC class I is scattered over 2000 kb, and the most centromeric class I gene is HLA-B and the most telomeric is HLA-F. To characterize the telomeric portion of the class I region, which is least characterized, I constructed four cosmid libraries from the total DNA of each of the yeast strains containing YACs 745D12, 960H11,

788E2 or 800G3 which is known to cover the class I region. Approximately 4500 transformants were picked from each library and screened with a radio-labeled human genomic DNA probe. The 348, 374, 95 and 187 clones were selected from YAC 745D12, 960H11, 788E2 and 800G3, respectively. The cosmids were further screened with 8 markers (S, HSRI, HLA-L, D6S265, HLA-G, HLA-F, MOG, and D6S131) and 2 cosmids (pM67 and pM78) in order to clarify their map locations. In addition, clones which contained the ends of each YAC could be identified by *URA3* and *TRP1* sequences present on the YAC vector. The cosmids could be ordered establishing 12 short contigs; and to fill the gaps, bi-directional chromosome walking was performed using an *Eco*RI fragment as the probe for each walking step (Fig. 10). Cosmid contigs were assembled by overlapping *Eco*RI bands and Southern blot analysis using the same probe as the chromosome walking (Fig. 11). *Eco*RI fragments found among many numbers of cosmids can ascertain the extent of overlap between the cosmids. Figure 12 shows the map of the 421 clones (TY-series cloned by me) and nine clones (p-series cloned by Tokai Univ.). Finally, these clones could be assembled into 8 contigs, which cover a total of 2.1 Mbp ranging from the telomeric portion of the class I region to its adjacent non-MHC region (Fig. 12). Because this *Eco*RI restriction technique generates maps in which the order of some internal fragments is not rigorously determined, technically the maps are not a standard *Eco*RI restriction map, but rather a *Eco*RI restriction fingerprint (Fig. 11A). However, by virtue of the large redundancy of fingerprinted clones that are used in their assembly, these maps are practically equivalent to standard restriction maps and are clearly suitable for clarifying the global GC% distribution in and around the MHC region.

Remaining gaps, except for the one between S and HSR-I gene, appeared difficult to be isolated by the present cosmid vector system. A size of the gap located between S and HSR-I has expected to be ca. 450-kb, and it becomes clear that the respective region is deleted in the YACs analyzed. Such mitotic instability of YACs was observed in other regions examined in this study; YAC 960H11 overlaps the two YACs containing HLA-F and MOG, but these two genes appeared only in the libraries derived from the latter two YACs. These gaps have recently

been cloned by the PAC-vector system by the Tokai Univ. group, and one of the PACs was used in this study (Fig. 12). Figures 9A and B show that a boundary of GC% mosaic domains exists in the junction between the MHC class I and its centromeric non-MHC region. I will explain later that this corresponds to a chromosome-band boundary, basing on the data of FISH on the human prometaphase chromosomes.

Along with the above chromosome walk in the telomeric portion of the class I, the walk in the centromeric portion adjoining the class III, was also conducted, because the preliminary analyses on the fragmental sequences indicated that the GC% levels of classes I and III are clearly different from each other. The walking was conducted as described above, and the contiguous cosmids obtained are presented by Fig. 12. Dr. Inoko and his colleagues in Tokai Univ. have conducted the large-scale sequencing of this class I region, and various cosmids which I isolated were used in the sequencing project as a collaborative work. The GC% distribution presented by Fig. 9A includes the result calculated from the sequences determined by Inoko's group. There appears to exist one GC% boundary in the junction between classes I and III. Again, the GC% boundary corresponds to a boundary of functional domains.

3. Chromosome in situ hybridization

The human MHC locus spans about 4 Mbp, exceeding the average size of bands observed by high-resolution banding. At a standard 850-band level the MHC locus is on a T-type R band 6p21.3. T-type R band (T bands) is an evidently heat-stable subgroup of R bands and correspond to GC-rich regions; they are called T bands because many of them are found at telomeric locations. By high-resolution banding, a thin G-positive sub-band, 6p21.32, was found within the MHC (Senger *et al.*, 1993). By *in situ* hybridization on metaphase chromosomes, the class II region was localized within sub-band 6p21.31 and the class I region to band 6p21.33 (Spring *et al.*, 1985).

In the non-MHC region telomeric to HLA-F (the most telomeric class I gene), anonymous markers such as RFP have been located by linkage analysis, and RFP was localized near the boundary between 6p21.3 and p22 (Vernet *et al.*, 1993). In the above chromosome walk

around the telomeric edge of class I, I also found both HLA-F and RFP. I further walked about 150 kb telomeric to RFP and found the region to be AT-rich (Figs. 9C and 12). By a standard fluorescence *in situ* hybridization (FISH) onto prometaphase, a cosmid TY6G11 located in 250 kb centromeric of RFP was mapped on 6p22.1 (G/Q) but close to 6p21.3 (R) (Figs. 9C and 13). This shows the walked region contains a boundary of chromosome bands. Very recently, Dr. Tenzen and I started to measure the replication timing in this region using the sets of PCR primers listed in Table 3. Confirming the previous results, two sites in MHC (AGER, and TNFA) replicate early, the first and/or second hour of S phase. In contrast, TY11B5T3 and TY7C10T3 sites (Fig. 9C) in the non-MHC region replicate late, the fifth and/or fourth hour of S phase (Fig. 14). This is consistent with the above finding that the AT-rich non-MHC region belongs to the G band 6p22.1 and the walked chromosome region actually contains a chromosome band boundary.

4. Characteristic structures in and around the GC% mosaic boundaries

The average size of human chromosome bands at the high resolution level has been estimated to be about 1.5 Mb (Comings, 1987). Considering the wide range of functional behaviors of chromosome bands and GC% mosaic domains, their boundaries are most probably composed of multiplex signals and structures ensuring multiple functions which may scatter in a region with a certain size. Based on the results obtained for the case of the classes II and III junction where the replication switching was clarified, we tentatively assumed the size to be several tens of kb. Concerning chromosome bands, at least three types (G/Q, ordinary R, T) were known for the human chromosomes, and concerning isochores, five types (L1, L2, H1, H2, H3) were known. Concerning replication timing, at least four types (very early, early, late, very late) were pointed out. When we consider the boundaries of GC% mosaic domains and of chromosome bands, various types are expected; e.g., boundaries between L1 and L2, those between L2 and H1, those between L2 and H2, and so on. I conducted the following types of

approaches to search and to list the characteristic structures in and around the boundary of GC% mosaic domains, which may relate to the possible functional signals and structures harbored by the band boundaries. One approach is to find the peculiar sequences which are easily recognizable; e.g. long polypurine/polypyrimidine sequences. The second is to find the human interspersed sequences with low or middle levels of frequencies. Two thousands of chromosome bands were visualized for the human chromosomes at the high resolution level, and the number of isochores were estimated to be 10^4. Considering the fact that several percent of human genome sequences were already registered by the International DNA Databases, the expected characteristic structures, if present, should have already been registered as the interspersed repetitive sequences such as MERs (medium reiteration frequency sequences). The third approach is to compare the human sequences having the GC% boundaries with the respective sequences of different organisms, on an assumption that the structures with functional significance should be conserved during evolution.

Long-range sequencing of the human genome was conducted by many groups and the sequences have been registered by DDBJ / EMBL / GenBank. By surveying the human genomic sequences longer than 100 kb, I found three genomic sequences (HS49J10, HSU91322, and HSU91318) which contain a boundary of the long-range GC% mosaic domains, and these sequences were included in the analyses. In the case of the HS49J10, two other sequences (HS267P14 and HS179I15A) can contiguously be linked at the ends, and its total sequence became 380 kb. The GC% distribution in each of these loci is shown by Fig. 15. For the first approach aforementioned which searches long polypurine / polypyrimidine sequences, AG% (A+G%) distribution was analyzed with a window size of 100 nt. The long polypurine/polypyrimidine sequences, which give the sharp high or low AG% peaks, are frequently observed in the GC% transition zones (Fig. 16). For the second approach to search the interspersed repeat sequences in the GC% transition zone, the 50-kb zone containing the most sharp GC% transition in each long sequence was analyzed by the RepeatMasker2

program and the found repeat sequences, other than *Alu* and LINE1 repeats, are presented in Fig. 15 and Table 4. Some MERs and MIRs common between the sequences are often found in the GC% transition zones. At the present moment, however, I can not find the sequence common among all GC% transition zones. This may be due to the fact that various type of GC% transitions were analyzed here; e.g. the transition between different L isochores found in HS49J10 and that between H and L isochores found in HUMMHCCLASSIII. For the third approach to search the evolutionary conserved sequences, I analyzed the human and mouse genomic sequences containing the junction between MHC classes II and III, which was recently registered by DDBJ / EMBL / GenBank. The GC% distribution of these human and mouse sequences is shown in Fig. 17. The evolutionary conserved sequences were searched using the dotter program (Sonnhammer et al., 1995), and about 1-kb sequence with significant homology between the human and mouse sequences was found in the GC% transition zone (Fig. 18). The 1 kb (Fig. 19) was neither a protein-coding sequence nor a simple repeat sequence. When DNA databases were searched using the 1-kb sequence as a query, partial similarity was found for several human sequences. At the present moment, further characteristic of this sequence is not known. It should be noted that there exists a long polypurine / polypyrimidine sequences also in this GC% transition zone of mouse.

3) Newly identified genes near the telomeric edge of the class I

In order to understand genome features of the walked MHC region from the functional and evolutionary views, the genes' distribution in the respective region was studied. For this purpose, at first, I determine the sequences of the terminal portion of the cloned fragments of individual cosmids, using the T3 or T7 promoter primer designed for the cosmid vector. Comparison of many terminal sequences (about 500 nt each) with GenBank data indicated evident similarities with various portions of the genes listed by Table 5. It should be stressed that because of the high density of the cosmids covering the individual regions, various portions of each gene sequence were found in a series of the consecutive cosmids. Confirming the proper walking, nine known genes having been mapped in the respective region were

identified; RFB30 (RING protein) gene, HLA-90, HLA-J, P5-1, MOG, diubiqutin gene, RCK-like gene, RFP, and $tRNA^{phe}$ in the direction from the class I to its neighboring non-MHC. The following nine genes were newly identified in the present study; SMT3B-like gene (designated as SMT3X in this study), GABAB receptor gene (GABR-B), mas-related gene, RASH-like gene, TRE17-like gene and MEA11-like gene, in the direction from the class I to the non-MHC, as well as three olfactory receptor genes. Interestingly, sequence similarity of these genes with those on chromosomes 7, 11, and 17, was disclosed, which will be discussed later. Among the newly identified genes in this region, I thought that GABR-B is the gene with the most interesting function, thus determined its complete sequence as explained below. Besides the genes above noted, at least eleven distinct genomic portions showed the evident homology to the ESTs registered by the International DNA Databases; seven of them had the identical sequences to the ESTs in the databases (Table 5).

4) Genomic organization of the human metabotropic gamma-aminobutyric acid receptor B (GABAB).

Gamma-aminobutyric acid (GABA) is the most abundant and widely distributed inhibitory neurotransmitter present in the central nervous system (CNS). The receptors activated with GABA have been classified into ionotropic (GABAA/GABAC) and metabotropic (GABAB) receptors. GABAA/GABAC receptors flow through chloride ions into the cell with GABA binding. These receptors belong to the diverse class of multi-subunit proteins that form 'ligand-gated' ion channels, though GABAB receptor is G-protein coupled receptor. While the precise number of GABAB receptor subtypes has not been determined, pharmacological results suggest that GABAB receptors can be subclassified into at least two types. One type of GABAB receptor is known to be sensitive to cis-aminocrotonic acid (CACA), but insensitive to baclofen which is a selective agonist for GABAB receptor (Matthews *et al.*, 1994). Concerning the molecular structure, only one GABAB receptor gene has been identified for rat, that was recently reported

by Kaupmann *et al.* (1997).

Since GABAB receptors modulate synaptic transmission by presynaptic inhibition of transmitter release or by increasing a K^+-conductance responsible for long-lasting inhibitory postsynaptic potentials (late IPSP), presynaptic and postsynaptic $GABA_B$ receptor subtypes was proposed (Bowery, 1993; Misgeld *et al.*, 1995; Dutar and Nicoll, 1988; Cunningham and Enna, 1996). Presynaptically, GABAB autoreceptors have been described controlling the release of GABA, whereas GABAB heteroreceptors regulate the release of L-glutamate, noradrenaline, dopamine, 5-hydroxytryptamine, substance P, cholecystokinin, and somatostatin. The induction of long-term potentiation, an associative increase in synaptic strength that may underlie the formation of some types of learning and memory, is affected by the activation of pre- and post-synaptic GABAB receptors. While the exact membrane topology of the GABAB receptor is unknown, the GABAB receptor is members of the 7TM (seven transmembrane) G-protein-coupled receptors which includes metabotropic glutamate receptors and Ca^{2+}-sensing receptors (Kaupmann *et al.*, 1997). G-protein which associates with GABAB receptors, contains Gi subunit (Nishikawa *et al.*, 1997), and GABAB receptor is known to negatively regulate adenylate cyclase level in mammalian cells, and negatively regulate transcription of the genes mediated by CRE binding protein (CREB) in CNS (Barthel *et al.*, 1996). It is unknown whether the expression of GABAB receptor itself is regulated via CRE and CREB. At least two pharmacologically distinct subclasses of GABAB receptor were shown to regulate formation of cyclic-AMP production in rat brain (Cunningham and Enna, 1996), and two types of rat cDNAs encoding GABAB receptor were identified (Kaupmann *et al.*, 1997). In the present study, I identified the human GABAB receptor gene near the telomeric edge of the MHC class I region, and determined the complete genomic organization of the GABAB receptor disclosing different promoters for the two isoforms, one of which has a CRE. This suggests that the activation of GABAB receptor controls relative expression of its two isoforms via the CRE- and CREB-mediated system.

1. Identification of GABAB receptor gene near the telomeric edge of the MHC class I and genomic sequence determination

As described above, I did chromosome walking around the telomeric edge of the class I using the cosmid libraries derived from for the four contiguous YACs (CEPH mega-YAC 745D12, 960H11, 788E2, and 800G3) known to cover the MHC class I region. During the walking and partial sequencing of the obtained cosmids, I identified 13 contiguous clones with high sequence homology to the rat GABAB receptor cDNA sequence recently reported by Kaupmann *et al.* (1997). The contig covered a physical distance of ca. 70 kb (Fig. 20) and contained the 5' portion of MOG gene which was mapped 60 kb telomeric apart from HLA-F by Pham-Dinh *et al.*, (1995). The GABAB receptor gene was thus located in the distal portion of MHC class I region, and this human gene was designated GABR-B in the present study. Another gene with high sequence homology with a human homologue of *Saccharomyces cerevisiae* SMT3 gene (Meluh and Koshland, 1995) was found upstream of the GABAB receptor gene.

In order to determine the complete genomic sequence of GABAB receptor, the cosmid TY11F3 was extensively characterized. Restriction enzyme fragments of TY11F3 were subcloned, and sequenced. The complete sequence has been deposited in DDBJ / EMBL / GenBank under Accession No. AB009982 and AB009983. Given the high nucleotide homology between the human and rat GABAB receptor sequence, I could deduce the complete gene organization of the GABAB receptor. The nucleotide identity for the protein-coding region was 91% (Fig. 21), and all splice donor and acceptor sites conformed to the GT/AG rule (Table 6) and matched with the consensus sequences around the GT/AG sites (Table 7; Balvay *et al.*, 1993; Ey *et al.*, 1993). The gene for the human GABAB receptor (GABR-B) spans ca. 31 kb and is composed of 23 exons (Fig. 22). In addition, two putative promoters were identified which will be discussed later. Sequences of the eighteen human expressed sequence tags (AA181004, AA348199, Z744106, H19658, H25821, H41556, H43286, H50397, H51356, M78726, N56175, R71844, R73356, R76485, T07518, T81224, T99207, W01458) which

were found by the homology search against dbEST database, were consistent with the predicted GABAB receptor gene sequence and enabled me to assign the entire 1.3-kb 3'-untranslated region (UTR) sequence with a polyadenylation signal, AATAAA, in the mRNA. This termination point was experimentally confirmed by the 3'RACE method.

The gene structure predicted by the GENSCAN program (Burge and Karlin, 1997) was almost identical to the present structure except for the short 21-bp exon 4. The probability of suboptimal exon structures was very low, indicating that different transcripts with other exons is unlikely to exist. Most of repetitive sequences, identified by RepeatMasker2 program (Smit and Green, http://ftp.genome.washinton.edu/cgi-bin/RepeatMasker) are clustered between exons 10 and 11 (Fig. 22). The dinucleotide repeat $(CA)_{22}$ is located 1 kb from the 5'-UTR, and showed VNTR polymorphism in different MHC haplotypes, which will be described later.

2. Two isoforms derived from alternate promoter usage

Based on the two distinct cDNA sequences for the rat GABAB receptor, Kaupmann *et al.* (1997) predicted two isoforms of the rat GABAB receptor; R1a of 961-amino acid residues and R1b of 845 residues. R1b differs from R1a in the N-terminal 163 residues which are replaced by 47 different residues. Utilizing the human genome sequences determined in the present study, complete sequences of the two human isoforms could be predicted. The human R1a isoform was one residue longer than the rat R1a, but the length of the R1b was identical. The identity level of amino acid sequence between the species was 98% for R1a and 97% for R1b (Fig 23). Aside from the glutamate and Ca^{2+} receptors previously reported, a homology search of the human GABAB receptor sequence using BLAST and FASTA showed similarity to the putative pheromone receptor of mouse (AF011411, AF011412, and AF011414) and rat (AF016181).

The CLUSTAL-W alignment program was used to multiple alignment of 15 mammalian 7TM G-protein-coupled receptors (for L-glutamate, Ca^{2+}, and pheromone) plus putative coding sequences of 7TM G-protein coupled receptors of *C. elegans* and *D. melanogaster,* which were

found by the sequence-homology search (*C. elegans*; U58748, *D. melanogaster*; AC002502). Figure 24 shows the alignment for the transmembrane domains of seventeen 7TM G-protein-coupled neurotransmitter receptors. Sequences in the third transmembrane domain and third intercellular loop (ICL) are significantly conserved with nine amino acids conserved among all the examined receptors. The conservation of cysteine residues found in the transmembrane domains suggests that these residues are important for the function and/or structure of the receptors. The cysteines in the sequence at the position of 188, 220, 663, and 761 in the human GABAB receptor (R1a) probably form disulfide bonds. Particularly, the cysteine residues at the position of 663 and 761 were located in the first and second extracellular loop (ECL), respectively (Fig. 24). Interestingly, such disulfide bonds and conservation of Cys residues between ECLs have been reported in muscarinic acetylcholine receptors whose primary sequences significantly differ from the GABA and glutamate receptors (Hulme *et al.*, 1990; Wess, 1993). The structural similarity of the 7TM G-protein-coupled receptors including both the GABAB receptor and the muscarinic acetylcholine receptor is intriguing phenomenon. This structural architecture of 7TM G-protein-coupled receptors may suggest that both family of 7TM G-protein-coupled receptors have diverged from a common origin.

To understand the mechanisms and factors that control GABAB receptor expression, clarification of the promoter region is crucial. As noted above, Kaupmann *et al.* (1997) found two mRNA forms of the rat GABAB receptor which encode two isoforms (R1a and R1b) and suggested that they may be produced by alternative splicing. The present sequence for the GABAB receptor is the first example of the genomic sequence of this gene, which will clarify the structures responsible for controlling its gene expression and the predicted alternative splicing. The alignment of the human genomic sequence with the rat cDNA sequence of the shorter isoform R1b, showed that the first exon of the R1b mRNA is present within the fifth intron of the long R1a isoform. Furthermore, the nucleotide sequence encoding the R1b-specific 47 amino acids was found to be directly connected with exon 5 of R1a isoform. I predicted the promoter for R1b isoform would be within the intron between exons 4 and 5 of

the R1a isoform and the two isoforms are derived from alternative promoter usage rather than alternative splicing (Fig. 25).

As a separate approach, putative promoters were searched by combining the results of the promoter search using the NNPP program with those of the search for the consensus transcription-factor binding motifs and for CpG islands. Two putative promoters with high scores were identified by the NNPP program and their positions were consistent with the gene structures for the R1a and R1b isoforms defined above, thus confirming that the promoter for the short isoform R1b is present in the intron between exons 4 and 5 of the long isoform R1a (Fig. 25). It should also be noted that the size of either mRNA generated by the alternative promoter usage is equivalent to the previously reported sizes for the rat mRNAs. As shown in Figs. 26A and 26B, each 5'-flanking genomic sequence immediately upstream of the initiation codon contained polypyrimidine/polypurine tract. Kirkness and Fraser (1993) reported that such pyrimidine rich tract play an important role in the transcriptional activity of the $GABA_A$ receptor ß3 subunit gene. Both the first and second promoter regions are GC rich (73% and 69% G+C) with 84 and 66 CpG di-nucleotides in 973-bp and 878-bp, respectively and are therefore, classified as CpG islands (Fig. 27). Such CpG islands are characteristically observed in the 5' portion of a wide range of genes, and the state of methylation of cytosine residues in CpG di-nucleotides is one of the mechanism that determines relative activity of promoter elements in these CpG islands (Bird, 1986). Though the methylation pattern of both CpG islands in the human $GABA_B$ gene is presently unknown, the methylation may control CNS specific gene expression in this case. Neither promoter contains CAAT or TATA motif. Such GC-rich and TATA-less promoters are known to be associated with ubiquitously expressed housekeeping gene (Dynan, 1986; Bird, 1986), and can be found upstream of several genes of the G-protein-coupled receptors (Collins *et al.*, 1993; Hauser *et al.*, 1994; Ilkuyama *et al.*, 1992; Wang *et al.*, 1992) and other tissue-specific genes such as synapsin-I, aldolase C, neural cell adhesion molecule, and olfactory neuron-specific proteins (Hirsch, 1990; Kudrycki, 1993; Makeh, 1994; Sauerwald, 1990).

The most significant difference between the two promoter regions is the presence of a canonical CRE in the second promoter region (Fig. 26B). The CRE has been identified in the promoter region of many cyclic AMP-inducible genes (Montminy *et al.*, 1990). Interestingly transcription of only one isoform of the GABAB receptor can be regulated by the CRE-binding protein, which will be discussed later. Other motifs are equally distributed in the both promoters. Putative recognition sequences for transcription factors AP2 and Sp1 are present 14 and 3 sites, respectively (Figs. 26A and 27). In addition, in the first promoter, the 15 putative binding sites for GC-factor (GCF) were found around the initiation codon (Figs. 26-A and 27); the GCF is transcriptional repressor that binds to GC-rich sequences (Kageyama and Pastan, 1989). The transcription from the first promoter may be affected by the transcription from the second promoter which is regulated by cAMP response element (Figs. 26B and 27).

3. (CA)n polymorphism found for different MHC haplotypes

The presence of the neurogenic gene in the human MHC region is of particular clinical interest because of the possible existence of genes responsible for several neurogenic diseases in this region. For example, dyslexia has been mapped on chromosome 6p21.3 (Grigorenko *et al.*, 1997; Cardon *et al.*, 1994). Polymorphic markers are valuable for investigating possible association of the present gene with such diseases. Di-nucleotide repeats $(CA)_{22}$ was found in the 1 kb upstream of the 5'-UTR of the GABAB gene. To investigate whether the (CA)n microsatellite repeats can be an informative polymorphic marker, a variable number of tandem repeats (VNTR) polymorphism were tested using 36 different Japanese HLA haplotype B-cell lines by PCR and successive sequencing (Table 8). The repeat length ranged from $(CA)_9$ to $(CA)_{24}$ giving 8 alleles.

4. Differentiation of two isoform's expression via alternative usage of promoters with and without the cyclic AMP-response element

The development of the CNS and its plasticity result from complex interactions within the neural network. At the cellular level, interactions mediated by neurotransmitters exert both acute

and long-term effects, and the latter may operate at the genomic level by switching genetic programs. At the genomic level, responses to extracellular signals are mediated by discrete regulatory sequences present in many genes. A well characterized target is the cyclic AMP-responsive element (CRE) which preferentially binds the transacting factor CREB (cyclic AMP-responsive element binding protein). This regulatory element, which mediates the transcriptional effect that is induced by the cyclic AMP-dependent regulatory pathway, is found in many neurogenic genes. The CRE sequence is also associated with the Ca^{2+}-dependent regulatory pathway. In the CNS, this CRE may represent a key target for many different hormones and neurotransmitters which are coupled with either the cyclic AMP or Ca^{2+} regulatory pathway. The present study shows that the CRE is present in the promoter region of one isoform of GABAB receptors (Figs. 26B and 27). GABAB receptors are known to affect short-term signaling in various cell types, but little is known about their possible long-term effects on gene expression. In cerebellar granular neurons, GABAB receptors are known to be coupled to the cyclic AMP- and Ca^{2+}-dependent regulatory pathways. Barthel *et al.* (1996) reported that GABAB receptors negatively regulate CREB-mediated transcription in the CNS. Since our findings show that only one isoform has the promoter with CRE, it suggest the GABAB receptor gene to be autoregulated on the transcription level via CRE in the second promoter; the activation of GABAB receptor gene may differentiate numbers and relative abundance of the two isoforms of the receptor (Fig. 28). This is consistent with the findings of Kaupmann *et al.* (1997) that the ratio of the isoforms is different in several rat brain tissues. Therefore, we hypothesize that, although an alternative promoter usage is known to be involved usually in tissue-specific expression of many genes (Simpson *et al.*, 1997), the promoters described here may response to the circumstantial changes rather than being tissue-specific.

In conclusion, I described the complete sequence of the human GABAB receptor gene. The gene, which encoded alternatively produced mRNA isoforms (R1a and R1b), spans 31 kb and consists of 23 exons. The both mRNAs are transcribed by TATA-less promoter and the 5'-

flanking region of both transcripts contain Sp1, AP2, and GCF binding motifs. In addition, the cAMP-responsive element is present only in the promoter for the R1b isoform. Comparison of amino acid sequences of the GABAB receptors with similar receptors suggests that all 7TM G-protein-coupled receptors have similar structural architecture. The present result has led to novel insights into the molecular mechanisms underlying the expression of the GABAB gene and the function of the GABAB receptor.

DISCUSSION

1) Band boundary and evolutionary process to produce segmental GC% distribution.

Chromosome bands are structures that can be observed through microscopes, and, when only this view is stressed, the precise location of their boundaries appears to be meaningless. However, considering various genome features having been connected with chromosomal band zones, band boundaries can presumably be assigned precisely by putting informative landmarks on genome DNA. The present study actually shows that the switching point for DNA replication timing can be fairly accurately assigned at the transition region of GC%. This also indicates that the long-range GC% mosaic is the structure with profound biological significance. If the long-range GC% mosaic is the structure which happened to have been formed without significant functions through evolution, the boundary may neither be sharp nor be precisely correlated with the replication timing because of the randomization due to a large number of neutral mutations accumulated in the course of evolution. Conspicuous correlation between the replication timing and the segmental GC% distribution indicates that the long-range GC% mosaic structure has biological functions and/or that the molecular mechanisms for evolutionary formation of the mosaic structure are still working. It is conceivable that the difference in replication timing have caused the segmental GC% distribution through certain molecular mechanisms during evolution (e.g. differential mutation rates and/or patterns related with replication timings), and some (if not all) of the evolutionary mechanisms may be still working. The present finding that all GC% boundaries in and around the human MHC correspond to the boundaries of the functional domains, supports the view that the GC% mosaic structures have biological functions. The strategy used in this study should be a general method for defining chromosome band boundaries at the molecular level and for putting valuable landmarks on mammalian genomes, which will help us to comprehensively understand molecular genetic and cytogenetic observations.

2) Linkage similarity of genes on 6p21.33 - 6p22.1 with those on chromosomes 7,11, and 17

As noted in INTRODUCTION, sets of genes which have sequence similarities with those on 6p21.3 were found in the restricted zones on chromosomes 1, 9 and 19 (Table 1). In the case for the newly found genes near the telomeric edge of class I, as well as those previously located in this region, sequence similarity was found for genes on chromosomes 7, 11, and 17, rather than those on chromosomes 1, 9 and 19 ; Table 9 shows six sets of genes with the sequence similarity. This shows a possible boundary of the genome multiplication during evolution and/or the genome rearrangement after the multiplication to be located near the telomeric edge of class I. Elucidation of correlation of this type of boundaries with those of GC% mosaic domains and of DNA replication timings, should give profound knowledge of evolutionary processes and mechanisms to establish the present human and mammalian genomes.

3) Possible association of GABR-B with genetic diseases

Susceptibility to a large number of diseases is thought to be associated with genes in the MHC region (Klein, 1986), and this genome portion has been intensively studied, disclosing various disease-related genes (Campbell and Trowsdale, 1993). As explained in RESULTS, GABAB receptors are known to be coupled to a variety of effector systems such as adenylate cyclase. Activation of the GABAB receptor exert inhibitory effects by interfering with the bioelectric activation of neurons. Therefore, a defect in this receptor may cause neurogenic diseases. In fact, GABAB-receptor antagonists affect cognition performance in rodents (Modadori, Moebius, and Zingg, 1996; Getova, Bowery, and Spassov, 1997). On human chromosome 6p21.3, on which the MHC is located, candidate genes responsible for neurogenic disease such as reading disability (Cardon et al., 1994; Grigorenko et al., 1997) have been located. In order to connect the present gene with such disease-responsible gene, I identified a polymorphic repeat sequence. The (CA)n length variation found in the 1 kb upstream of the 5'-UTR of the

GABAB gene can be used as a marker to show an association with human genetic diseases. The heterozygosity value (> 0.3) of this repeat locus indicates that the (CA)n repeat will be useful as an informative polymorphic marker.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. J. Mol. Biol. **215**, 403-410.

Aota, S., and Ikemura, T. (1986). Diversity in G+C content at the third position of codons in vertebrate genes and its cause. Nucl. Acids Res. 14, 6345-6355 & 8702 (for erratum).

Balvay, L., Libri, D., and Fiszman, M. Y. (1993). Pre-mRNA secondary structure and the regulation of splicing. BioEssays **15**, 165-169.

Baran, N., Lapidot, A., and Manor, H. (1991). Formation of DNA triplexes accounts for arrests of DNA synthesis at d(TC)n and d(GA)n tracts. Proc. Natl. Acad. Sci. USA **88**, 507-511.

Barthel, F., Kienlen Campard, P., Demeneix, B. A., Feltz, P., and Loeffler, J. ph. (1996). GABAB receptors negatively regulate transcription in cerebellar granular neurons through cyclic AMP responsive element binding protein-dependent mechanisms. Neuroscience **70**, 417-427.

Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M., and Rodier, F. (1985). The mosaic genome of warm-blooded vertebrates. Science **228**, 953-958.

Bernardi, G. (1989). The isochore organization of the human genome. Annu. Rev. Genet. **23**, 637-661.

Bernardi, G. (1993). The isochore organization of the human genome and its evolutionary history - a review. Gene **135**, 57-66.

Bettecken, T., Aissani, B., Muller, C. R., and Bernardi, G. (1992). Compositional mapping of the human dystrophin- encoding gene. Gene **122**, 329-325.

Bird, A. P. (1986). CpG-rich islands and the function of DNA methylation. Nature **321**, 209-213.

Bird, A. P. (1987). CpG islands as gene markers in the vertebrate nucleus. Trends Genet. **3**, 342-347.

Brinton, B. T., Caddle, M. S., and Heintz, N. H. (1991). Position and orientation-dependent effects of a eukaryotic Z-triplex DNA motif on episomal DNA replication in COS-7 cells. J. Biol. Chem. **266**, 5153-5161.

Bowery, N. G. (1993) GABAB receptor phamacology. Annu. Rev. Phamacol. Toxicol. **33**,109-147.

Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. **268**, 78-94.

Burt, M. J., Smit, D. J., Pyper, W. R., Powell, L. W., Jazwinska, E. C. (1995). A 4.5-megabase YAC contig and physical map over the hemochromatosis gene region. Genomics **33**, 153-158.

Campbell, R. D., and Trowsdale, J. (1993). Map of the human MHC. Immunol. Today **14**, 349-352.

Cardon, L. R., Smith, S. D., Fulker, D. W., Kimberling, W. J., Pennington, B. F., and DeFries, J. C. (1994). Quantitative trait locus for reading disability on chromosome 6. Science **266**, 276-279.

Collins, S., Ostrowski, J., and Lefkowitz, R.J. (1993). Cloning and sequence analysis of the human beta 1-adrenergic receptor 5'-flanking promoter region. Biochim Biophys Acta **1172**, 171-174.

Coming, D. E. (1978). Mechanisms of chromosome banding and implications for chromosome structure. Annu. Rev. Genet. **12**, 25-46.

Contreas, G., Giacca, M., and Falaschi, A. (1992). Purification of BrdUrd-substituted DNA by immunoaffinity chromatography with anti-BrdUrd antibodies. BioTechniques **12**, 824-825.

Craig, J. M., and Bickmore, W. A. (1994). The distribution of CpG islands in mammalian

chromosomes. Nature Genet. **7**, 376-382.

Cunningham, M. D., and Enna, S. J. (1996). Evidence for pharmacologically distinct GABAB receptors associated with cAMP production in rat brain. Brain Res. **720**, 220-224.

Diviacco, S., Norio, P., Zentilin, L., Menzo, S., Clementi, M., Biamonti, G., Riva, S., Falaschi, A., and Giacca, M. (1992). A novel procedure for quantitative polymerase chain reaction by coamplification of competitive templates. Gene **122**, 313-320.

Drouin, R., Holmquist, G. P., Richer, C. L. (1994). High-resolution replication bands compared with morphologic G- and R-bands. Adv Hum Genet **22**, 47-115.

Dutar, P. and Nicoll, R.A.A. (1988). A physiological role for GABAB receptors in the central nervous system. Nature **332**, 156-158.

Ellis, N., and Goodfellow, P. N. (1989a). The mammalian pseudoautosomal region. Trends Genet. **5**, 406-410.

Ellis, N. A., Goodfellow, P. J., Pym, B., Smith, M., Palmer, M., Frischauf, A.-M., and Goodfellow, P. N. (1989b). The pseudoautosomal boundary in man is defined by an Alu repeat sequence inserted on the Y chromosome. Nature **337**, 81-84.

Ellis, N., Yen, P., Neiswanger, K., Shapiro, L. J., and Goodfellow, P. N. (1990). Evolution of the pseudoautosomal boundary in Old World monkeys and great apes. Cell **63**, 977-986.

Ey, P. L., Datby, J. M., Andrews, R. H., and Mayrhofer, G. (1993). Giardia intestinalis:detection of major genotypes by restriction analysis of gene amplification products. Int. J. Parasitol. **5**, 591-600.

Falaschi, A., Giacca, M., Zentilin, L., Norio, P., Diviacco, S., Dimitrova, D., Kumar, S., Tuteja, R., Biamonti, G., Perini, G., Weighardt, F., Brito, J., and Riva, S. (1993). Searching for replication origins in mammalian DNA. Gene **135**, 125-135.

Fukagawa, T., Sugaya, K., Matsumoto, K., Okumura, K., Ando, A., Inoko, H., and Ikemura, T. (1995). A boundary of long-range G+C% mosaic domains in the human MHC locus: pseudoautosomal boundary-like sequence exists near the boundary.

Genomics **25**, 184-191.

Fukagawa, T., Nakamura, Y., Okumura, K., Nogami, M, Ando, A., Inoko, H., Saitou, N., and Ikemura, T. (1996). Human pseudoautosomal boundary-like sequences: expression and involvement in evolutionary formation of the present-day pseudoautosomal boundary of human sex chromosomes. Hum. Mol. Genet. **5**, 23-32.

Gardiner, K., Aissani, B., and Bernardi, G. (1990). A compositional map of human chromosome 21. EMBO J. **9**, 1853-1858.

Getova, D., Bowery, N. G., Spassov, V. (1997) Effects of GABAB receptor antagonists on learning and memory retention in a rat model of absence epilepsy. European Journal of Pharmacology **320**, 9-13.

Grigorenko, E., L., Wood, F. B., Meyer, M. S., Hart, L. A., Speed, W. C. Shuster, A., and Pauls, D. L. (1997). Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15. Am. J. Hum. Genet. **60**, 27-39.

Hauser, F., Meyerhof, W., Wulfsen, I., Schonrock, C., and Richter, D. (1994). Sequence analysis of the promoter region of the rat somatostatin receptor subtype 1 gene. FEBS Lett. **345**, 225-228.

Hidaka, M., Akiyama, M., and Horiuchi, T. (1988). A consensus sequence of three DNA replication terminus sites on the E. coli chromosome is highly homologous to the terR sites of the R6K plasmid. Cell **55**, 467-475.

Hirsch, M. R., Gaugler, L., Deagostini-Bazin, H., Bally-Cuif, L., and Goridis, C. (1990). Identification of positive and negative regulatory elements governing cell-type-specific expression of the neural cell adhesion molecule gene. Mol. Cell. Biol. **10**, 1959-1968.

Holmquist, G., Gray, M., Porter, T., and Jordan, J. (1982). Characterization of Giemsa dark- and light-band DNA. Cell **31**, 121-129.

Holmquist, G. P. (1987). Role of replication time in the control of tissue specific gene

expression. Am. J. Hum. Genet. **40**, 151-173.

Huberman, J. A., and Riggs, A. D. (1968). On the mechanism of DNA replication in mammalian chromosomes. J. Mol. Biol. **32**, 327-341.

Hulme, E. C., Birdsall, N. J. M., and Buckley, N. J. (1990). Muscarinic receptor subtypes. Annu. Rev. Pharmacol. Texicol. **30**, 633-673.

Ikemura, T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. **2**, 13-34.

Ikemura, T., and Aota, S. (1988). Global variation in G+C content along vertebrate genome DNA: Possible correlation with chromosome band structures. J. Mol. Biol. **203**, 1-13.

Ikemura, T., Wada, K., and Aota, S. (1990). Giant G+C% mosaic structures of the human genome found by arrangement of GenBank human DNA sequences according to genetic positions. Genomics **8**, 207-216.

Ikemura, T., and Wada, K. (1991). Evident diversity of codon usage patterns of human genes with respect to chromosome banding patterns and chromosome numbers; relation between nucleotide sequence data and cytogenetic data. Nucleic Acids Res. **19**, 4333-4339.

Ikuyama, S., Niller, H.H., Shimura, H., Akamizu, T., and Kohn, L.D. (1992). Characterization of the 5'-flanking region of the rat thyrotropin receptor gene. Mol. Endocrinol. **6**, 793-804.

Jurka, J., Kaplan, D. J., Duncan, C. H., Walichiewicz, J., Milosavljevic, A., Gayathri, M., and Solus, J. F. (1993). Identification and characterization of new human medium reiteration frequency repeats. Nucleic Acids Res. **21**, 1273-1279.

Kageyama, R., and Pastan, I. (1989). Molecular cloning and characterization of a human DNA binding factor that represses transcription. Cell **59**, 815-825.

Kasahara, M., Hayashi, M., Tanaka, K., Inoko, H., Sugaya, K., Ikemura, T., and Ishibashi, T. (1996). Chromosomal localization of the proteasome Z subunit gene reveals an ancient chromosomal duplication involving the major histocompatibility

complex. Proc. Natl. Acad. Sci. USA **93**, 9096-9101.

Kaupmann, K., Huggel, K., Heid, J., Flor, P. J., Bischoff, S., Mickel, S. J., McMaster, G., Angst, C., Bittiger, H., froestl, W., and Bettler, B. (1997). Expressin cloning of GABAB receptors uncovers similarity to metabotropic glutamate receptors. Nature **386**, 239-246.

Katsanis, N., Fitzgibbon, J., and Fisher, E. M. C. (1996). Paralogy mapping: identification of a region in the human MHC triplicated onto human chromosomes 1 and 9 allows the prediction and isolation of novel PBX and NOTCH loci. Genomics 35, 101-108.

Kirkness, E. F. and Fraser, C. M. (1993). A strong promoter element is located between alternative exons of a gene encoding the human g-aminobutyric acid-type A receptor b3 subunit (GABRB3). J. Biol.Chem. **286**, 4420-4428.

Klein, J. (1986). Natural History of the Major Histocompatibility Complex, Wiley, New York.

Kobayashi, T., Hidaka, M., Nishizawa, M., and Horiuchi, T. (1992). Identification of a site required for DNA replication fork blocking activity in the rRNA gene cluster in Saccharomyces cerevisiae. Mol. Gen. Genet. **233**, 355-362.

Korenberg, J. R., and Rykowski, M. C. (1988). Human genome organization: Alu, Lines and the molecular structure of metaphase chromosome bands. Cell **53**, 391-400.

Kudrycki, K., Stein-Izsak, C., Behn, C., Grillo, M., Akeson, R., and Margolis, F. (1993). Olf-1-binding site: Characterization of an olfactory neuron-specific promoter motif. Mol. cell. Biol. 13, 5, 3002-3014.

Little, R. D., Platt, T. H. K., and Schildkraut, C. L. (1993). Initiation and termination of DNA replication in human rRNA genes. Mol. Cell. Biol. **13**, 6600-6613.

Makeh, I., Thomas, M., Hardelin, J. P., Briand, P., Kahn, A., and Skala, H. (1994). Analysis of a brain-specific isozyme. J. Biol. Chem. **269**, 4194-4200.

Matsumoto, K., Arai, M., Ishihara, N., Ando, A., Inoko, H., and Ikemura, T. (1992a).

Cluster of fibronectin type III repeats found in the human major histocompatibility complex class III region shows the highest homology with the repeats in an extracellular matrix protein, Tenascin. Genomics **12**, 485-491.

Matsumoto, K., Ishihara, N., Ando, A., Inoko, H., and Ikemura, T. (1992b). Extracellular matrix protein tenascin-like gene found in human MHC class III region. Immunogenetics **36**, 400-403.

Matthews, G., Ayoub, G. S., Heidelberger, R. (1994). Presynaptic inhibition by GABA is mediated via two distinct GABA receptors with novel pharmacology. J. Neurosci. **14**, 1079-1090.

Meluh, P. B., Koshland, D. (1995). Evidence that the MIF2 gene of Saccharomyces cerevisiae encodes acentromere protein with homology to the mammalian centromere protein CENP-C. Mol. Biol. Cell. **6**, 793-807.

Misgeld, U., Bijak, M., and Jarolimek, W. (1995). A physiological role for GABAB receptors and the effects of baclofen in the mammalian central nervous system. Prog. Neurobiol. **46**, 423-462.

Modadori, C., Moebius, H-J., and Zingg, M. (1996) CPG36742, an orally active GABAB receptor antagonist, facilitates memory in a social recognition test in rats. Behavioural Brain reserch **77**, 227-229.

Montminy, M. R., Gonzalez, G. A. and Yamamoto, K. K. (1990) regulation of cAMP-inducible genes by CREB. trends. neurosci. **13**, 184-188.

Mouchiroud, D., D'Onofrio, G., Aissani, B., Macaya, G., Gautier, C., Bernardi, G. (1991). The distribution of genes in the human genome. Gene **100**, 181-187

Nishikawa, M., Hirouchi, M., and Kuriyama, K. (1997). Functional coupling of Gi subtype with GABAB receptor/adenylyl cyclase system: analysis using a reconstituted system with purified GTB-binding protein from bovine cerebral cortex. Neurochem. Int. **31**, 21-25.

Pearson, W. R., and Lipman, D. J. (1988). Improved tools for biological sequence

comparison. Proc. Natl. Acad. Sci. USA **85**, 2444-2448.

Pedrali-Noy, G., Spadari, S., Miller-Faurés, A., Miller, A. O. A., Kruppa, J., and Koch, G. (1980). Synchronization of HeLa cell cultures by inhibition of DNA polymerase a with aphidicolin. Nucleic Acids Res. **8**, 377-387.

Pham-Dinh, D., Jones, E. P., Pitiot, G., Della Gaspera, B., Daubas, P., Mallet, J., Le Paslier, D., Lindahi, F. K., and Dautigny, A. (1995). Physical mapping of the human and mouse MOG gene at the distal end of the MHC class Ib region. Immunogenetics **42**, 386-391.

Pilia, G., Little, R. D., Aissani, B., Bernardi, G., and Schlessinger, D. (1993). Isochores and CpG islands in YAC contigs in human Xq26.1-qter. Genomics **17**, 456-462.

Pilz, A., Prohaska, R., Peters, J., and Abbott, C. (1994). Genetic linkage analysis of the Ak1,

Saccone, S., De Sario, A., Wiegant, J., Raap, A. K., Della Valle, G., and Bernardi, G. (1993). Correlation between isochores and chromosomal bands in the human genome. Proc. Natl. Acad. Sci. USA **90**, 11929-11933.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, 2nd ed., (Cold Spring Harbor, New YorkL: Cold Spring Harbor Laboratory Press).

Sauerwald, A., Hoesche, C., Oschwald, R., and Kilimann, M. W. (1990). The 5'-klanking region of the synapsin I gene. J. Biol. Chem. **265**, 14932-14937.

Sealey, P. G., Whittaker, P. A., and Southern, E. M. (1985). Removal of repeated sequences from hybridisation probes. Nucl. Acids. Res. **13**, 1905-1922.

Senger, G., Ragoussis, J., Trowsdale, J., and Sheer, D. (1993). Fine mapping of the human MHC class II region within chromosome band 6p21 and evaluation of probe ordering using interphase fluorescence in situ hybridization. Cytogenet. Cell Genet. **64**, 49-53.

Siebert, P. D., and Larrick, J. W. (1992). Competitive PCR. Nature 359, 557-558.

Simpson, R. E., Dodson, M., Michael, V. R., Veena, R. A., M. M., Hinshelwood, Bulun, E. E., and Zhao, Y. (1997). Expression of the CYP19 (aromatase) gene: an unusual case

of alternative promoter usage. FASEB J, **11**, 29-36.

Sonnhammer, E. L., Durbin, R. (1995). A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene **167**, GC1-GC10.

Spring, B., Fonatsch, C., Muller, C., Pawelec, G., Kompf, J. Wernet, P., and Ziegler, A. (1985). Refinement of HLA gene mapping with induced B-cell line mutants. Immunogenet. **21**, 277-291.

Sugaya, K., Fukagawa, T., Matsumoto, K., Mita, K., Takahashi, E., Ando, A., Inoko, H., and Ikemura, T. (1994). Three genes in the human MHC class III region near the junction with the class II: gene for receptor of advanced glycosylation end products, PBX2 homeobox gene and a Notch homolog, human counterpart of mouse mammary tumor gene int-3. Genomics. **23**, 408-419.

Sugaya, K., Sasanuma, S., Nohata, J., Kimura, T., Fukagawa, T., Nakamura, Y., Ando, A., Inoko, H., Ikemura, T., and Mita, K. (1997). Gene organization of human NOTCH4 and (CTG)n polymorphism in this human counterpart gene of mouse proto-oncogene Int3. Gene **189**, 235-244.

Tenzen, T., Yamagata, T., Fukagawa, T., Sugaya, K., Ando, A., Inoko, H., Gojobori, T., Fujiyama, A., Okumura, K., and Ikemura, T. (1997). Precise switching of DNA replication timing in the GC content transition area in the human major histocompatibility complex. Mol. Cell. Biol. **17**, 4043-4050.

Therman, E. (1986). Human chromosomes: structure, behavior, effects. (New York: Springer-Verlag).

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl. Acids Res. **22**, 4673-4680.

Tribioli, C., Biamonti, G., Giacca, M., Colonna, M., Riva, S., and Falaschi. A. (1987). Characterization of human DNA sequences synthesized at the onset of S-phase. Nucleic

Acids Res. **15**, 10211-10232.

Wang, H.,Nelson, S., Ascoli, M., and Segaloff, D. L. (1992). The 5'-flanking region of the rat luteinizing hormone/chorionic gonadotropin receptor gene confers Leydig cell expression and negative regulation of gene transcription by 3',5'-cyclic adenosine monophosphate. Mol. Endocrinol. **6**, 320-326.

Wells, R. D., Collier, D. A., Hanvey, J. C., Shimizu, M., Wohlrab, F. (1988). The chemistry and biology of unusual DNA structures adopted by oligopurine.oligopyrimidine sequences. FASEB J. **2**, 2939-2949.

Wess, J. (1993). Molecular basis of muscarinic acetylcholine receptor function. Trends. pharmacol. Sci. **14**, 308-313.

Whitfield, L. S., Hawkins, T. L., Goodfellow, P. N. and Sulston, J. (1995) 41 kilobases of analyzed sequence from the pseudoautosomal and sex-determining region of short arm of human Y chromosome. Genomics, **27**, 306-311.

Vernet, C., Boretto, J., Mattei, M. G., Takahashi, M., Jack, L. J., Mather, I. H., Rouquier, S., Pontarotti, P. (1993). Evolutionary study of multigenic families mapping close to the human MHC class I region. J. Mol. Evol. **37**, 600-612.

Yoon, Y., Sanchez, J. A., Brun, C., and Huberman, J. A. (1995). Mapping of replication initiation sites in human ribosomal DNA by nascent-strand abundance analysis. Mol. Cell. Biol. **15**, 2482 2489.

LEGENDS TO FIGURES 1-2

Figure 1. Loci examined for replication timing in MHC classes II and III. Numbers indicate the position in the 450-kb region, spanning from HLA-DRA to TNX by kilobases (kb). TNFA is 500 and 870 kb away from TNX and HLA-DRA, respectively.

Figure 2. Competitive PCR products electrophoresed through an 8% polyacrylamide gel. The constant ^{3}H-dpm of BrdUrd-substituted DNA (usually 100-300 dpm in a tube) was co-amplified with constant amounts of competitor DNA (usually 1000-5000 molecules in a tube), which are 17 nt longer than the respective genomic sequences. The proper PCR conditions such as the annealing temperature and the ratio between nascent and competitor DNA were determined in advance. Competitive PCR reactions under a single condition optimized for each individual locus, were conducted three times (Fig. 3); the results of one series of experiments are shown here. The lanes represent 0-1 hour (lane 1), 1-2 hours (lane 2), 2-3 hours (lane 3), 3-4 hours (lane 4), 4-5 hours (lane 5), and 5-6 hours (lane 6). Bands from the competitor DNA and the BrdUrd-substituted nascent DNA are indicated by hollow and solid arrowheads, respectively. Bands of weaker intensity appeared at adjoining time intervals, presumably due to (1) incomplete cell synchronization (although it was synchronized twice), or (2) genuine cell-to-cell variation in replication timing. Faint bands at other time intervals are presumably due to the imperfect separation of BrdUrd-labeled DNA from unlabeled DNA.

Figure 1.

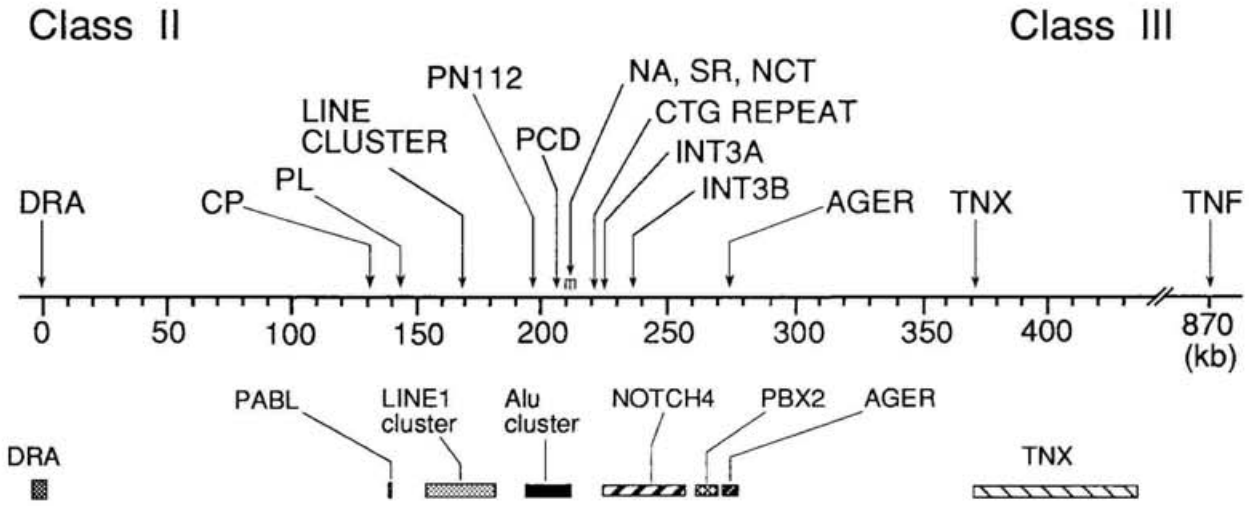
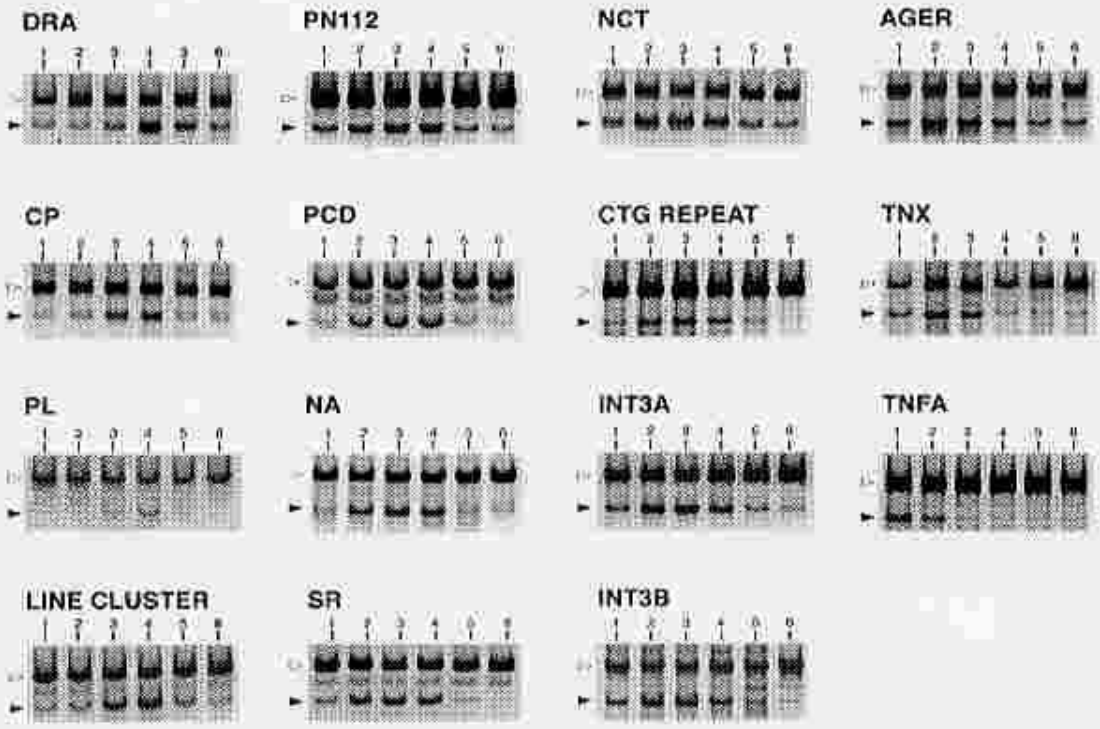

Class II
Class III
PN112
NA, SR, NCT
LINE
CLUSTER
CTG REPEAT
PCD
INT3A
PL
INT3B
DRA
CP
AGER
TNX
TNF
0
50
100
150
200
250
300
350
400
870
(kb)
PABL
LINE1
cluster
Alu
cluster
NOTCH4
PBX2
AGER
DRA
TNX

Figure 2.


DRA
PN112
NCT
AGER
CP
PCD
CTG REPEAT
TNX
PL
NA
INT3A
TNFA
LINE CLUSTER
SR
INT3B

LEGENDS TO FIGURES 3-4

Figure 3. Quantitative analysis of nascent DNA. The molecule number of nascent DNA containing each locus in 1 dpm of DNA was calculated by comparison with the intensity of the competitor band. The X-axis represents the labeling intervals and the Y-axis represents the number of molecules in 1 dpm of nascent DNA. Graphs represent the mean values calculated from three sets of independent experiments. The highest peak for all loci appeared at the same time period in independent PCR experiments. Each bar in the graph represents the maximum and minimum value.

Figure 4. Correlation between replication timing and GC% distribution in human MHC classes II and III. A) Replication timing. The Y-axis represents the replication timing; the time when S-phase was started by removing aphidicolin from the medium was taken as 0. The interval corresponding to the highest peak is assigned as the replication timing of the respective locus and the midpoint of the period is plotted. In the cases of PN112, PCD, INT3A, and INT3B, however, the differences between the highest and the second highest peaks were so small (mostly within a 10% difference, Fig. 3) that their midpoint is plotted. The overlapped dots situated about 210 kb from DRA correspond to the three closely located sites; NA, SR, and NCT (see Fig. 1). B) GC% distribution in the junction area between MHC classes II and III is from Fukagawa *et al.* (1995), and the GC% transition around the *Alu* cluster was designated "L/H transition".

Figure 3.

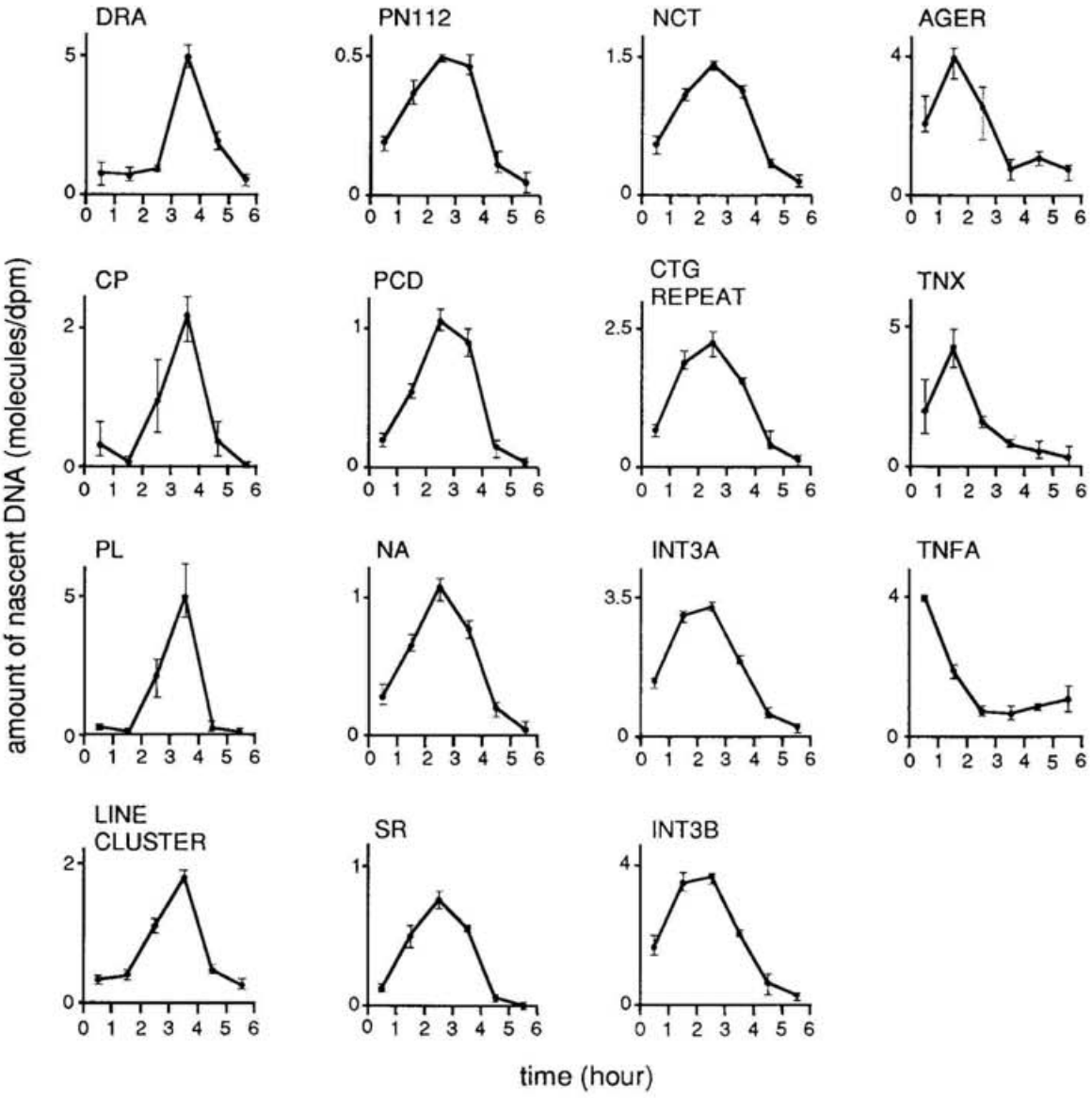
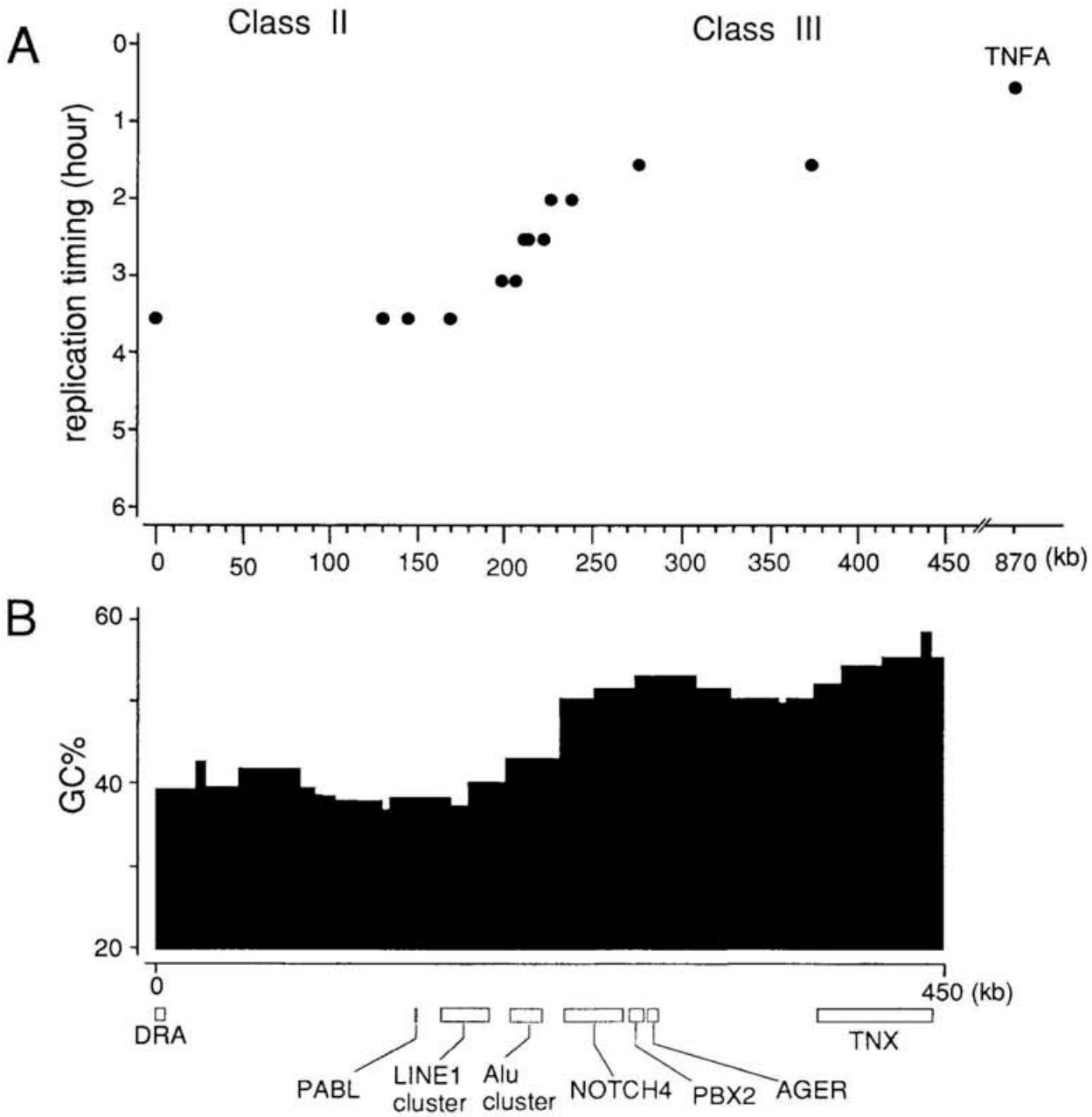

DRA
PN112
NCT
AGER
CP
PCD
CTG
REPEAT
TNX
PL
NA
INT3A
TNFA
LINE
CLUSTER
SR
INT3B
amount of nascent DNA (molecules/dpm)
time (hour)

Figure 4.


A
Class II
Class III
TNFA
replication timing (hour)
0
1
2
3
4
5
6
0
50
100
150
200
250
300
350
400
450
870 (kb)
B
GC%
60
40
20
0
450 (kb)
DRA
PABL
LINE1 cluster
Alu cluster
NOTCH4
PBX2
AGER
TNX

LEGENDS TO FIGURES 5-7

Figure 5. The 210-bp polypurine / polypyrimidine tract found at the exact mid point of the switching region of replication timing during S phase. A) Polypurine sequence of the 210 bp tract; from 5301 to 5510 nucleotide in GenBank Locus HSMHC3A5 (Accession No. U89335). B) An example of H-DNA triplexes formed for the 210-bp pur/pyr tract when only the perfectly consecutive Hoogsteen-pairing was assumed.

Figure 6. Characteristic features of the 5.7-kb region. A) Purine / pyrimidine ratio. Window size is 100 nt. B) Characteristic sequences. Solid arrows indicate the position and direction of *Alu* elements and four polypurine/polypyrimidine tracts are shown. Bars written as NA, SR and NCT indicate regions amplified by PCR. C) The 462-bp AT-rich (71%) sequence between SR and NCT. A/T tracts have been boxed. The position of this sequence is shown by the bracket in (B) and the 5'-end is proximal to SR.

Figure 7. Di-, tri-, or tetranucleotide short tandem repeats and MERs in and around the switch region. MERks3, PN112, and MERks2 are the new medium reiteration frequency sequences (MERs). MERks3 showed homology with a STS of human X chromosome (GenBank HUMSWX270) and an intron of retinoblastoma susceptibility gene (GenBank HUMRETBLAS). 1.7-kb PN112 showed homology with two human ESTs (expressed sequence tag), EST05686 (EMBL HS7963) and EST00737 (EMBL HSXT00737). 270-bp MERks2 showed homology with the 5' flanking region of CD22 gene (EMBL S61408).

Figure 5.

A

CGAAAGAAGG AAAGAAAGAA AGAAGGAAGG AAGGAAGGAA GGAAGGAAGG
AAAGAAGGAA GGAAGGAAGG AAGGAAAGAA GGAAGGAAGG AAAGAAAGAA
GGAAGGAAGG AAGGAAGGAA AGAAGGAATG AAGGAAGGAA GGAAGGAAAG
AAAGAAGGAA GGAAGGAAGG AAGGAAGGAA GGAAAGAAAG AAAGAAAGAA
AGAAAGAAAA

B

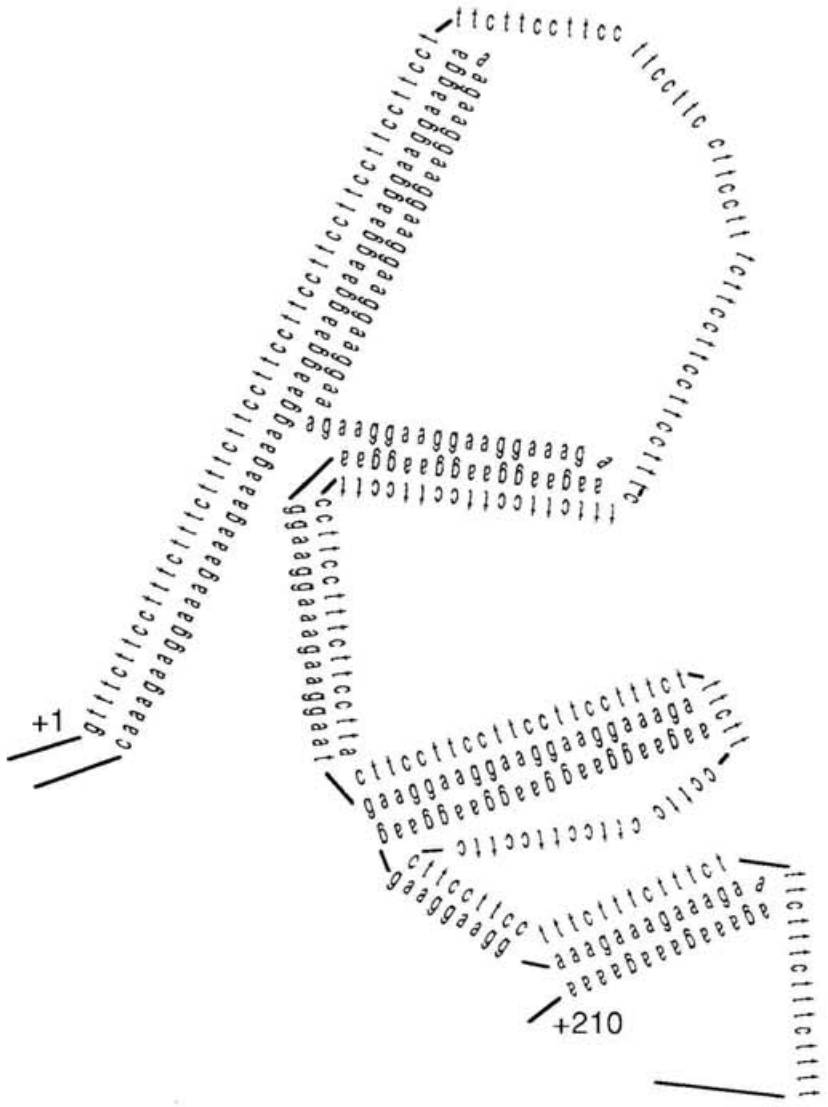
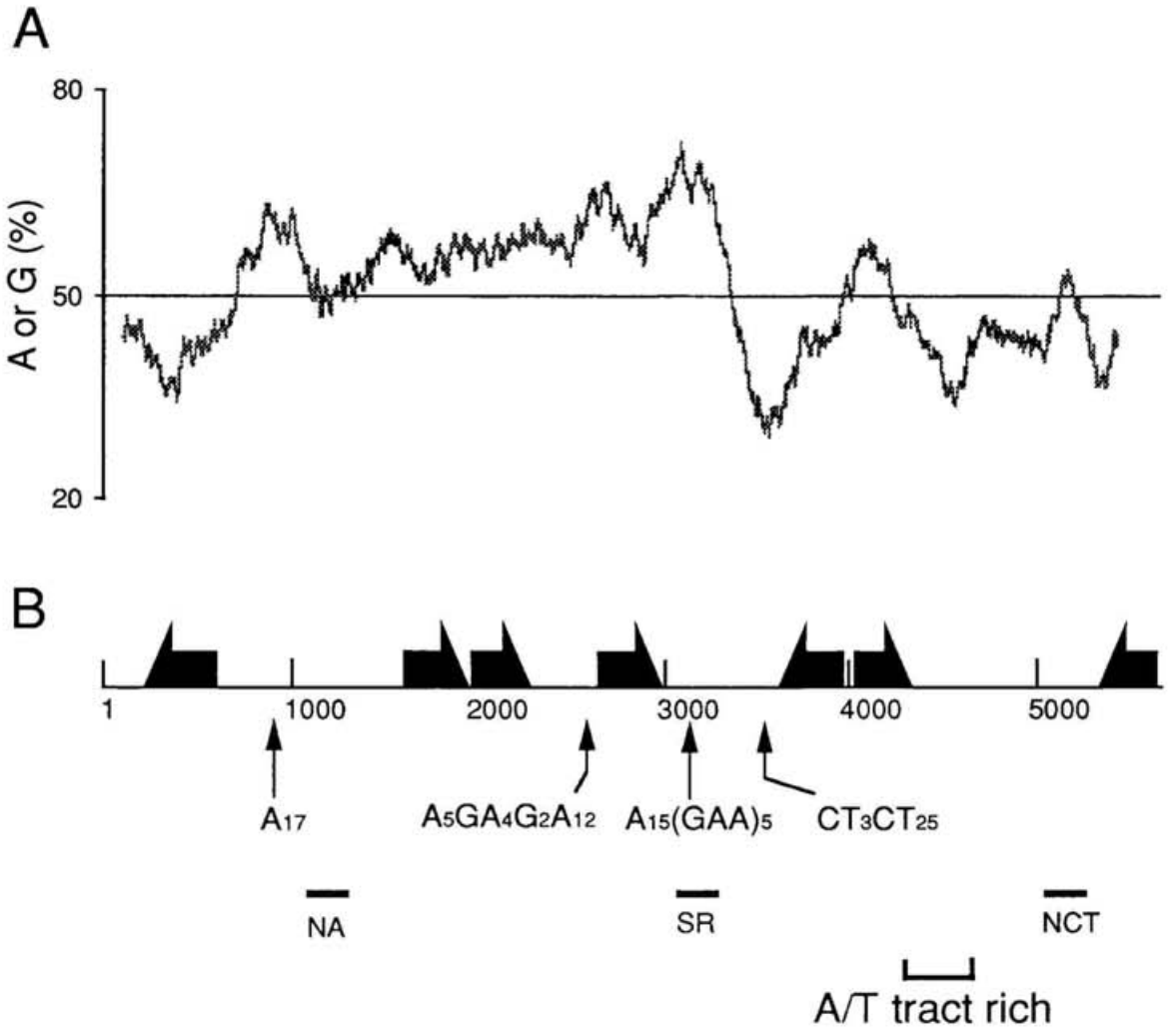


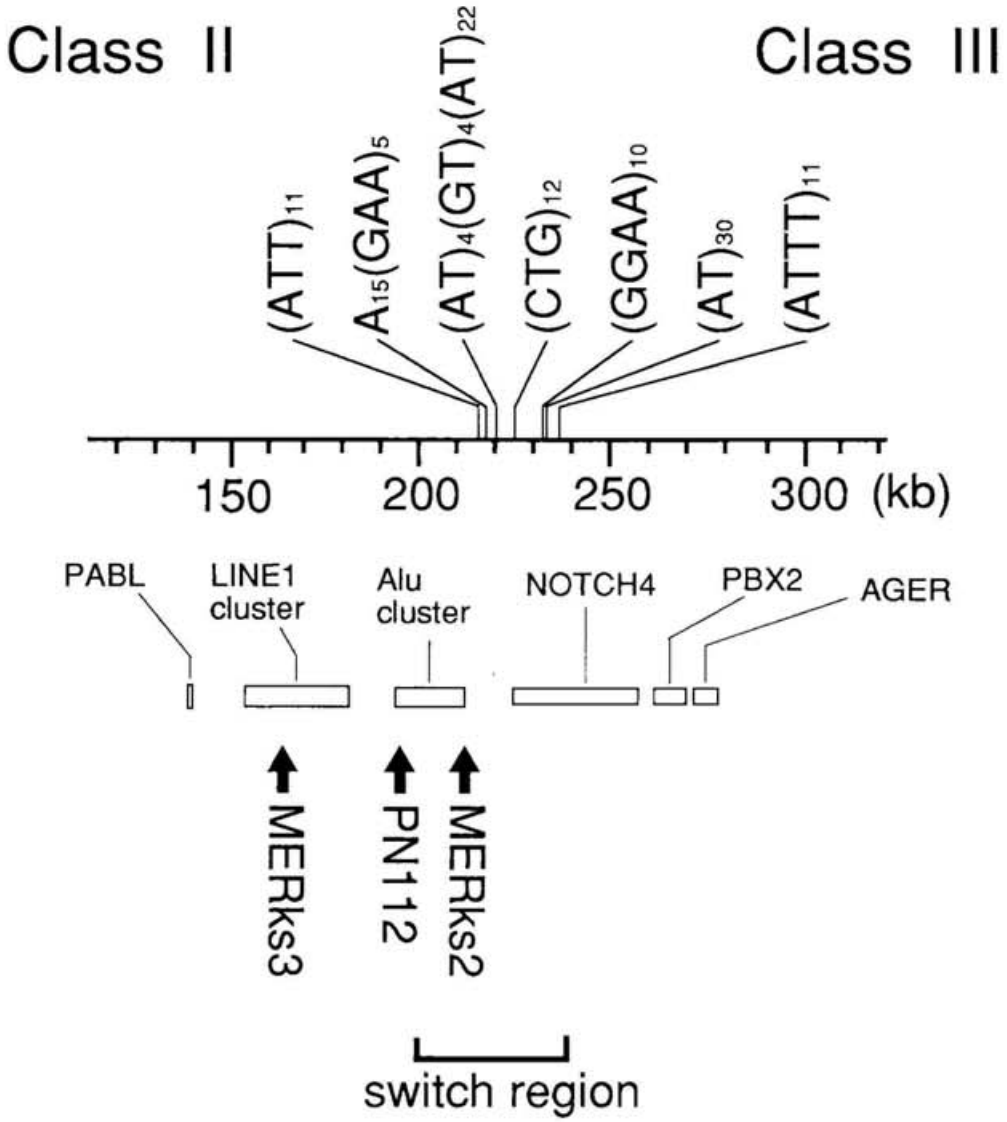
Figure 6.



C

CTATTTTATTGCTTGTGTAGATGTAAGAAAGTATCATATGTAAGAAATCATTGCCTAGCCCAACAT
CATGAATATCTACTTTTAGGCCTTTTTCTAATAGTTTTATAGTTTTGCACTTACATTTAGGTCTAT
GATCAATTTTAAGTTGGTTTTTTGTGTATGGTATGAGGTAATATAGGAATTACAAGCTATAAATCC
CCTCTAAGCATAGCTTTAGCTACATGTCAAAAATTAATTCTGATATGTTCTGTTTCAATTTTCATT
CAGCTTGAAATATTTTCTGGTTTTCCTTGTGATTTCTTCTTCCACCCATTGGTTGTTTAGAAGCAT
GCTGTTTAATTTTCACACACTTGTAATTTTCTCAAATTTCCTGTTGTTGTTGATTTCTAATTTAAT
TCTATTGTGATCAGAGAACATCTTTGTATTATTTCAGTCCTTTTAAAATTATTGAGACGTGTTTTA

Figure 7.


Class II
Class III
(ATT)11
A15(GAA)5
(AT)4(GT)4(AT)22
(CTG)12
(GGAA)10
(AT)30
(ATTT)11
150
200
250
300 (kb)
PABL
LINE1 cluster
Alu cluster
NOTCH4
PBX2
AGER
MERKs3
PN112
MERKs2
switch region

LEGENDS TO FIGURES 8-9

Figure 8. Physical map of contiguous cosmid clones that cover the junction between the MHC class II and its adjacent non-MHC. Names outlined by the box represent genes in this region. Position of restriction sites and fragment sizes for *Eco*RI are indicated in the map at the top of the figure. Dotted lines indicated gap between contigs. Ordered clones are represented by horizontal lines with vertical bars indicating *Eco*RI sites;the cosmid clones of AK series are indicated by thicker horizontal lines. DNA fragments, which were used to isolated the cosmid clones (AK series) as probes, are indicated by gray boxes. Clones used for sequencing are from the cosmid AK1G5.

Figure 9. A) Base-compositional map of the human MHC. Genomic MHC sequences in GenBank longer than 5 kb were selected, and their GC% was arranged by genetic position. GC contents of cloned fragments are measured directly by the biochemical method described under Materials and Methods. The sequences longer than 50 kb were divided into a span of 50 kb. Gene-encoding regions are known to be often GC-richer than their flanks. This produces local GC% fluctuations within an isochore and a tendency for thinner bars (usually corresponding to gene sequences) to be GC-richer than thicker bars (corresponding to both genes and their long flanks). B) Base-compositional map of the human MHC class II and its adjacent non-MHC region. Arrows indicate the positions of genes. Range of GC contents between HLA-DRA and HLA-DPA is roughly 40-42% level. The local high GC% of ca. 45% is observed two region; one containing TAP1, LMP7, TAP2 and HLA-DOB gene, and another containing RING3 and HLA-DMA. A GC% boundary was located in the junction area between the MHC class II and its adjacent non-MHC region. C) Base-compositional map of the MHC class I and its adjacent non-MHC region. Arrows indicate the positions of the cosmid clones, TY11B5, TY7C10, and TY6G1, as well as genes. The GC% level of the MHC class I region (from HSRI to HLA-F) is about 45%, but those of the non-MHC region are mostly less than 40%. There exists a GC% boundary in the junction area between the MHC class I and its adjacent non-MHC region.

Figure 8.

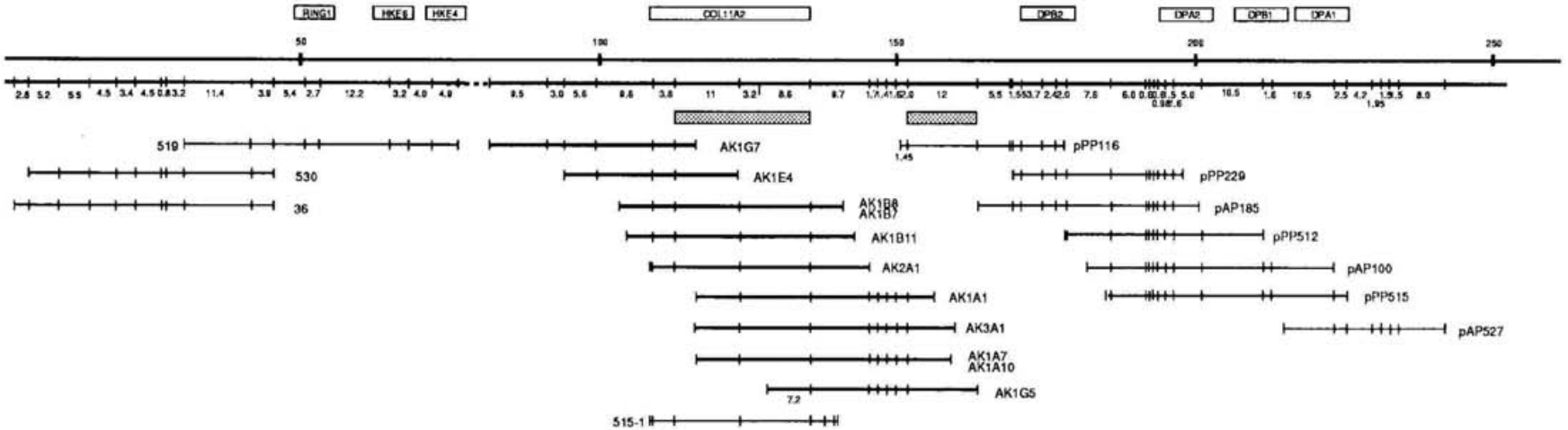

RING1
HKE6
HKE4
COL11A2
DPB2
DPA2
DPB1
DPA1
50
100
150
200
250
519
530
36
AK1G7
AK1E4
AK1B8
AK1B7
AK1B11
AK2A1
AK1A1
AK3A1
AK1A7
AK1A10
AK1G5
515-1
pPP116
pPP229
pAP185
pPP512
pAP100
pPP515
pAP527

Figure 9A.

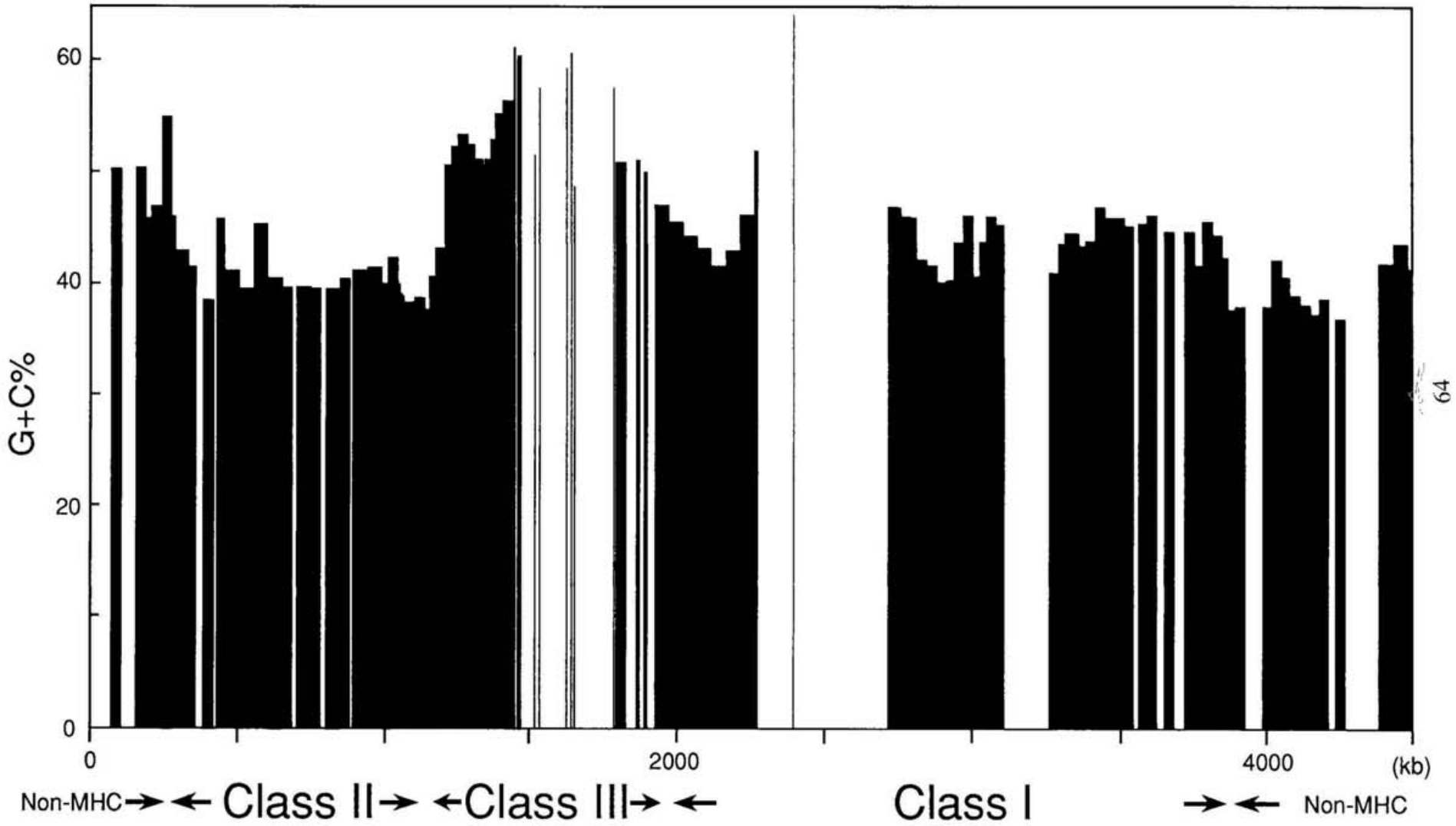

60
40
20
0
G+C%
0
2000
4000
(kb)
Non-MHC
Class II
Class III
Class I
Non-MHC

Figure 9B.

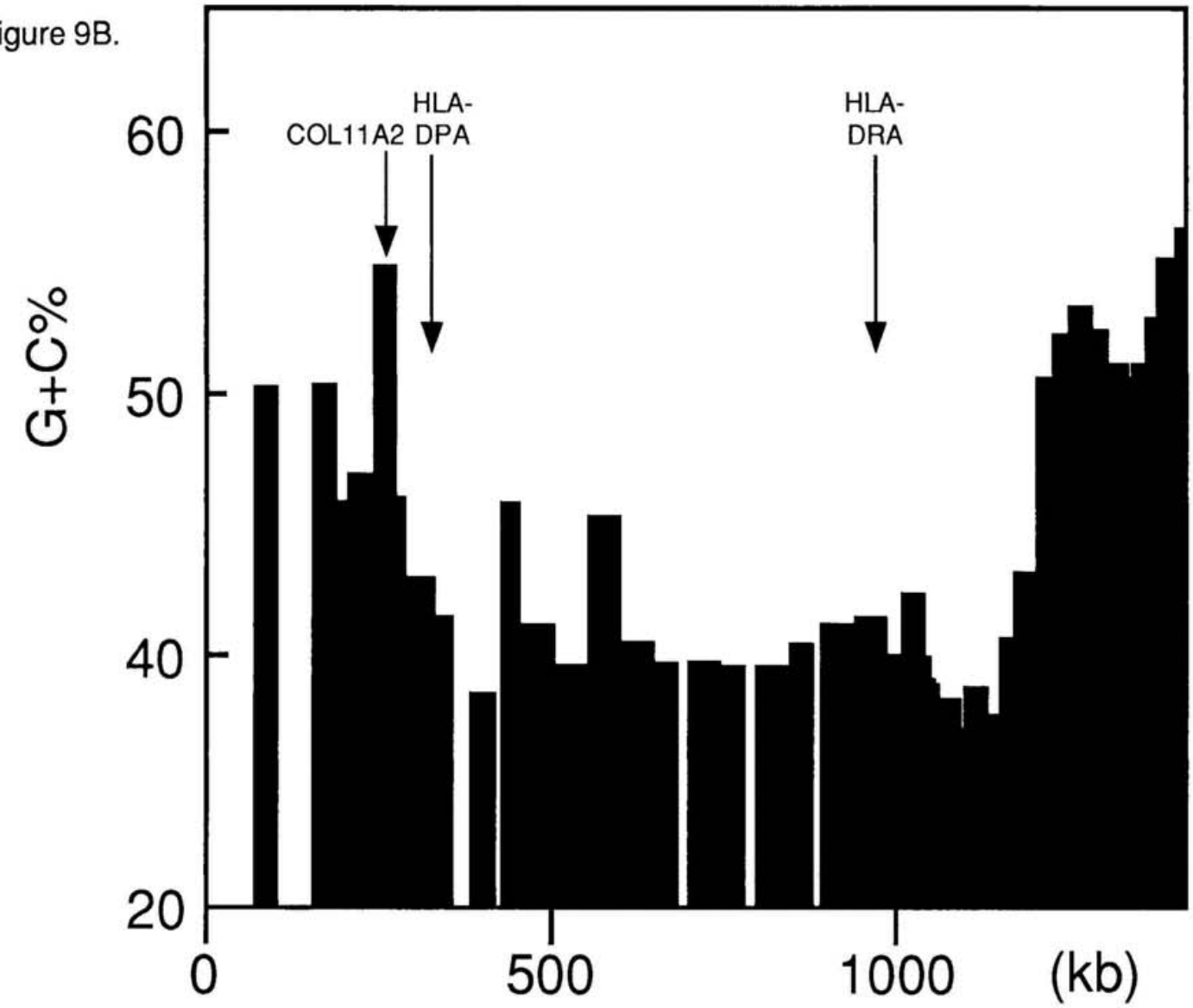
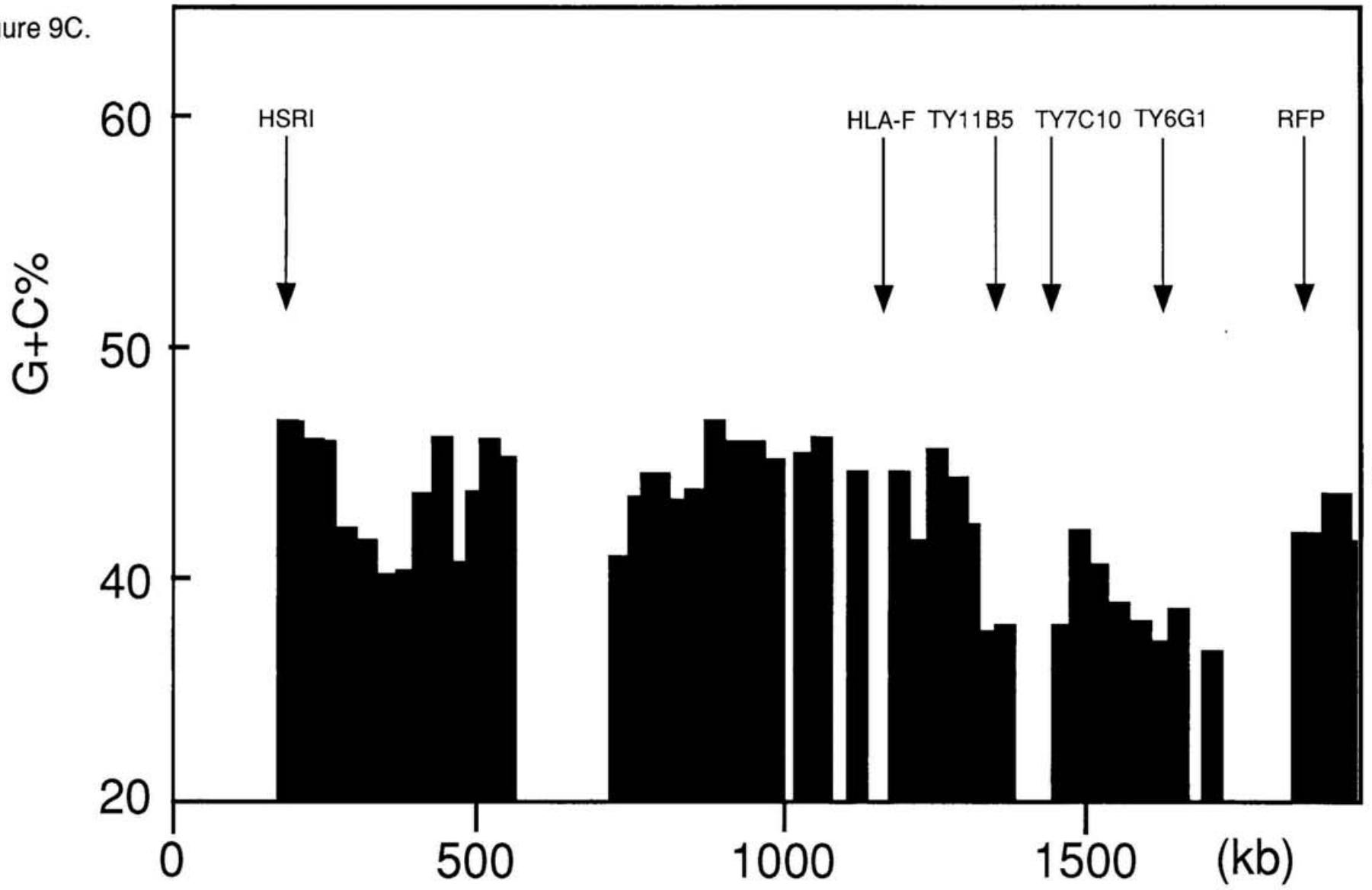


Figure 9C.


HSRI
HLA-F
TY11B5
TY7C10
TY6G1
RFP
G+C%
60
50
40
20
0
500
1000
1500
(kb)

LEGENDS TO FIGURES 10-13

Figure 10. An example of the chromosome walking by colony hybridization. Each clone was spotted onto a nylon membrane. *Eco*RI fragment of the cosmid clone (TY9E10) indicated by an arrowhead, was used as the probe. Thirteen clones gave positive signals.

Figure 11. Confirmation of overlaps between cosmids. A) The isolated cosmids were assembled into contigs by *Eco*RI-digestion profiles on an agarose gel. B) Fragments which did not gave the fragments of the same size, corresponded to the ends of insert DNAs, and indicated by triangles. Southern blot analysis was done with the same probe with the walking. In this analysis, there exists a false positive clone, TY9D10.

Figure 12. Physical map of contiguous cosmid and PAC clones that cover the junction between the MHC class I and its adjacent non-MHC region. Names outlined by the box represent genes in this region. Position of *Eco*RI restriction sites and the fragment sizes are indicated in the map at the top of the figure. Dotted lines indicate the gap between contigs, and DNA fragments used as probes are indicated by closed boxes. Ordered clones are represented by horizontal lines with vertical bars indicating *Eco*RI sites and the clone names are shown on the side. PAC clones are indicated by thicker horizontal lines. More than 70% of MHC class I region is covered with these clones.

Figure 13. Fluorescence *in situ* hybridization of a genomic clone onto human prometaphase chromosomes. TY6G11 was located on a chromosome band 6p22.1; arrows indicate signals. These results were based on observations made on more than ten prometaphase chromosomes.

Figure 10.

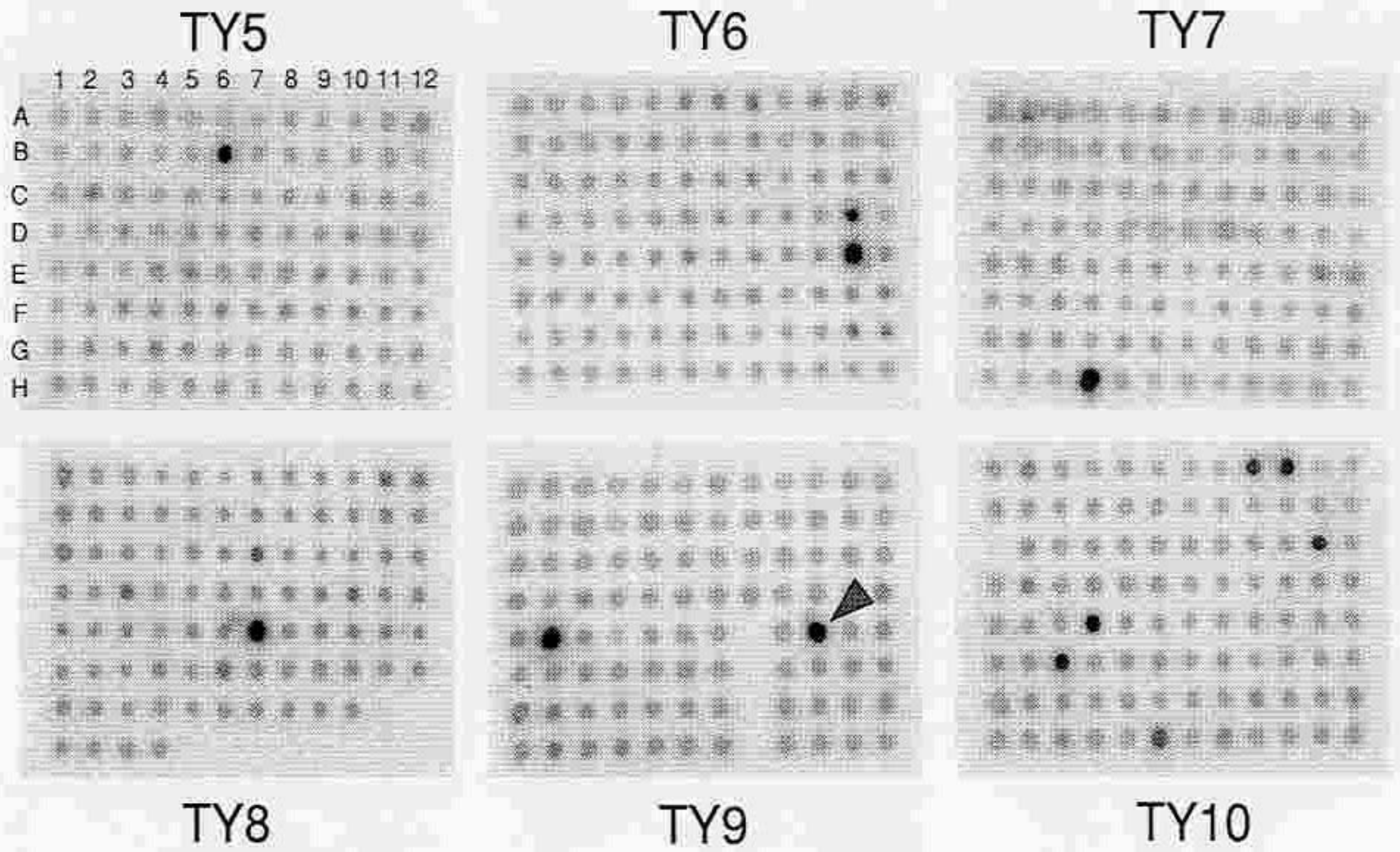


Figure 11.

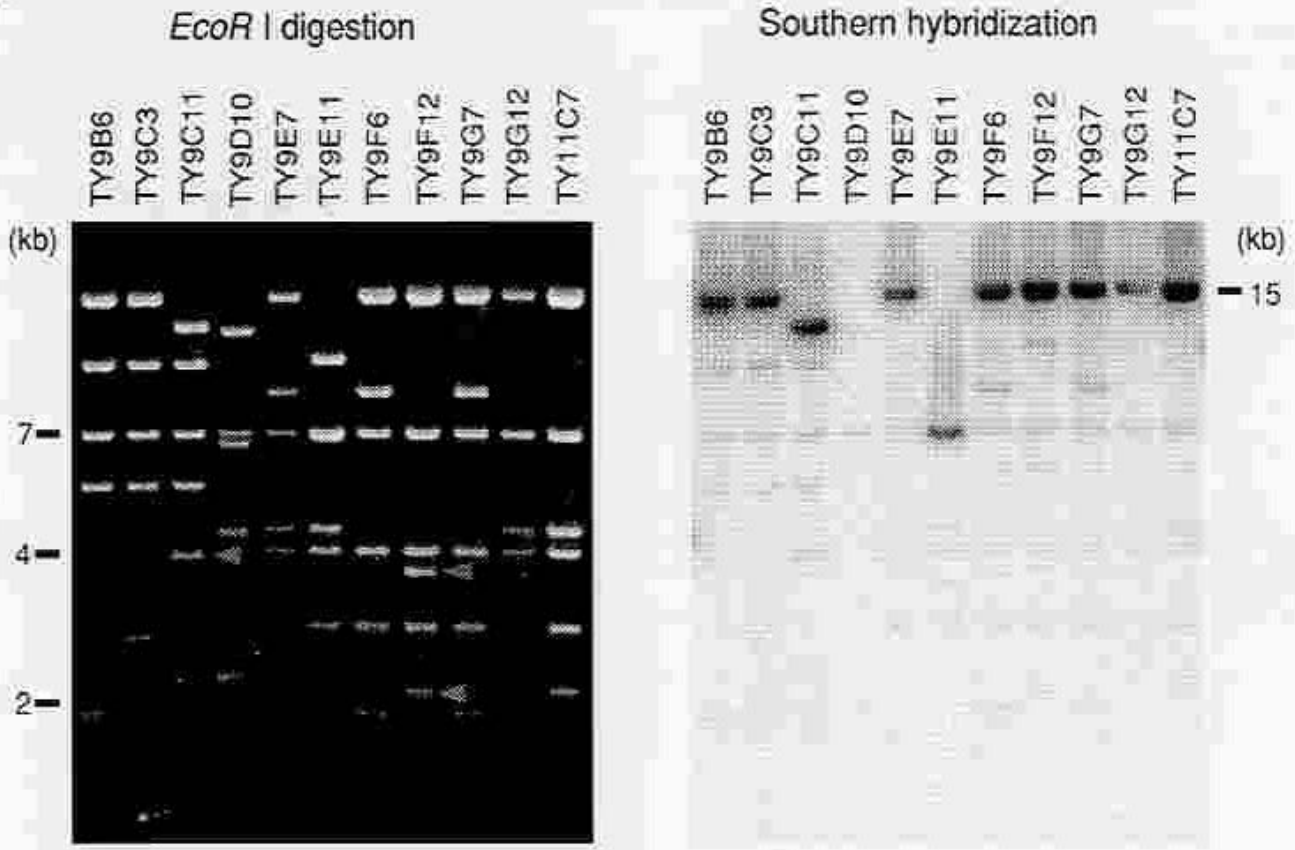

EcoR I digestion
Southern hybridization
TY9B6
TY9C3
TY9C11
TY9D10
TY9E7
TY9E11
TY9F6
TY9F12
TY9G7
TY9G12
TY11C7
(kb)
7
4
2
(kb)
15

Figure 12.

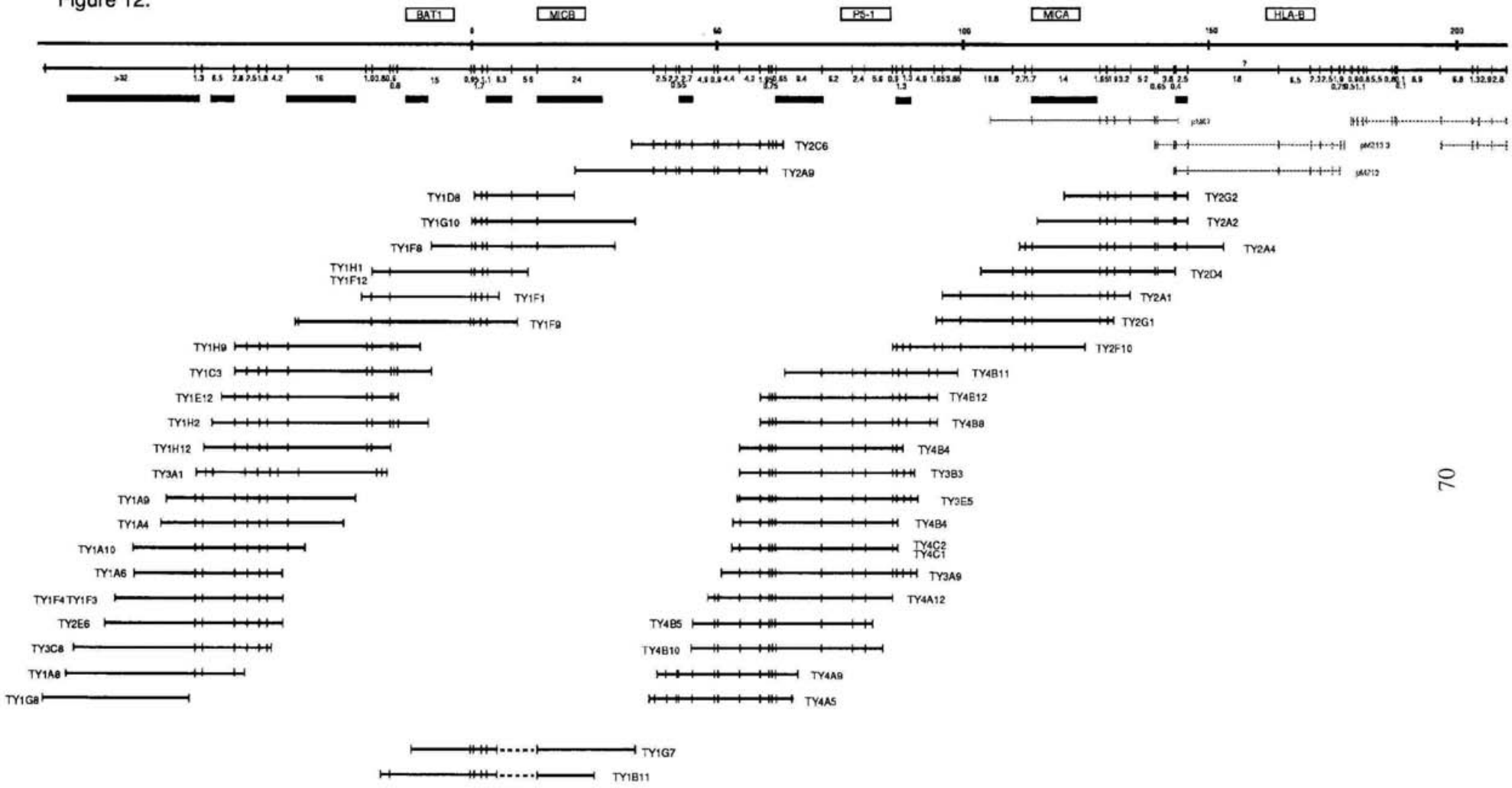

BAT1
MICB
PS-1
MICA
HLA-B
TY2C6
TY2A9
TY1D8
TY1G10
TY1F8
TY1H1
TY1F12
TY1F1
TY1F9
TY1H9
TY1C3
TY1E12
TY1H2
TY1H12
TY3A1
TY1A9
TY1A4
TY1A10
TY1A6
TY1F4 TY1F3
TY2E6
TY3C8
TY1A8
TY1G8
TY2G2
TY2A2
TY2A4
TY2D4
TY2A1
TY2G1
TY2F10
TY4B11
TY4B12
TY4B8
TY4B4
TY3B3
TY3E5
TY4B4
TY4C2
TY4C1
TY3A9
TY4A12
TY4B5
TY4B10
TY4A9
TY4A5
TY1G7
TY1B11

Figure 12. Continued

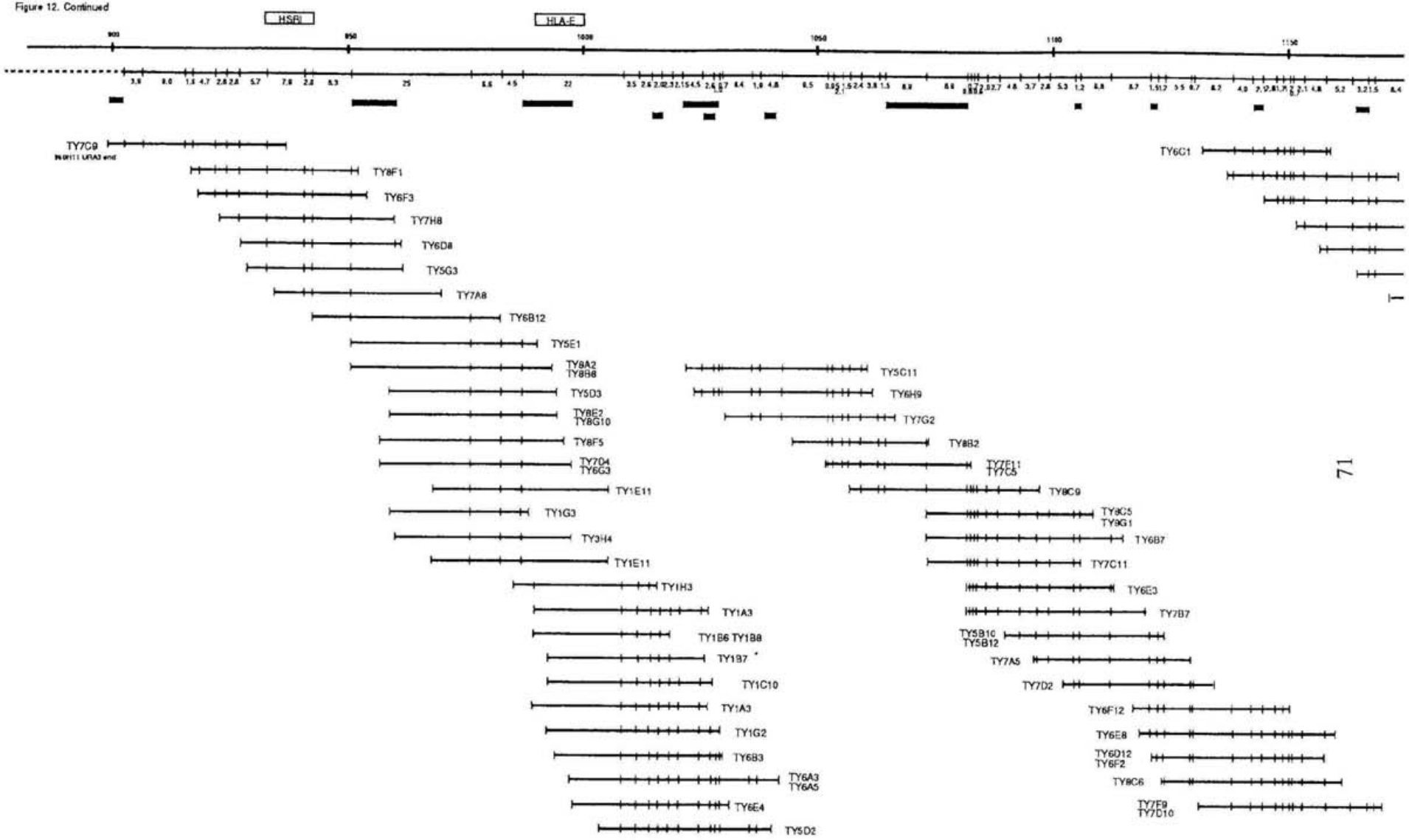


Figure 12. Continued

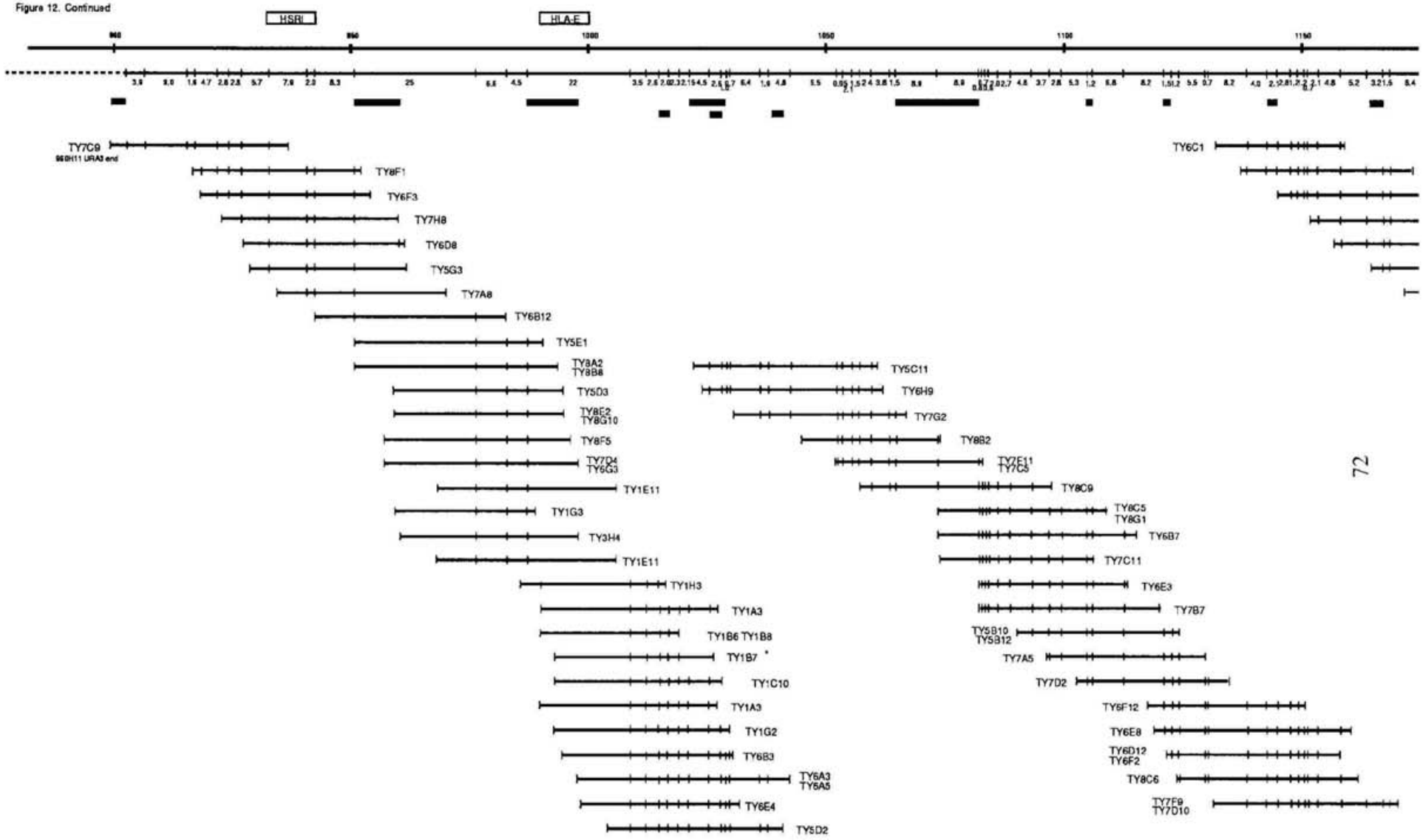


Figure 12. Continued

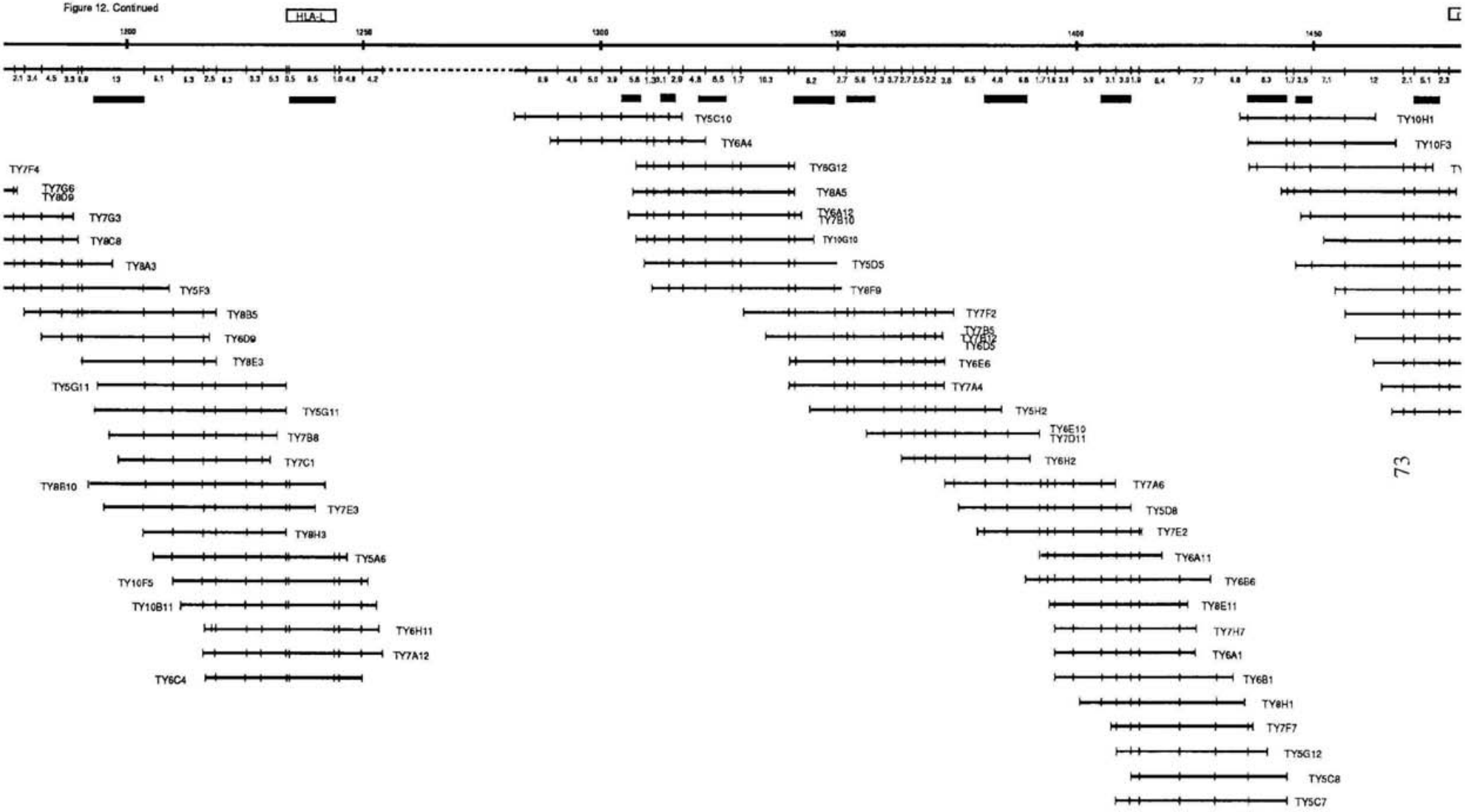


Figure 12. Continued

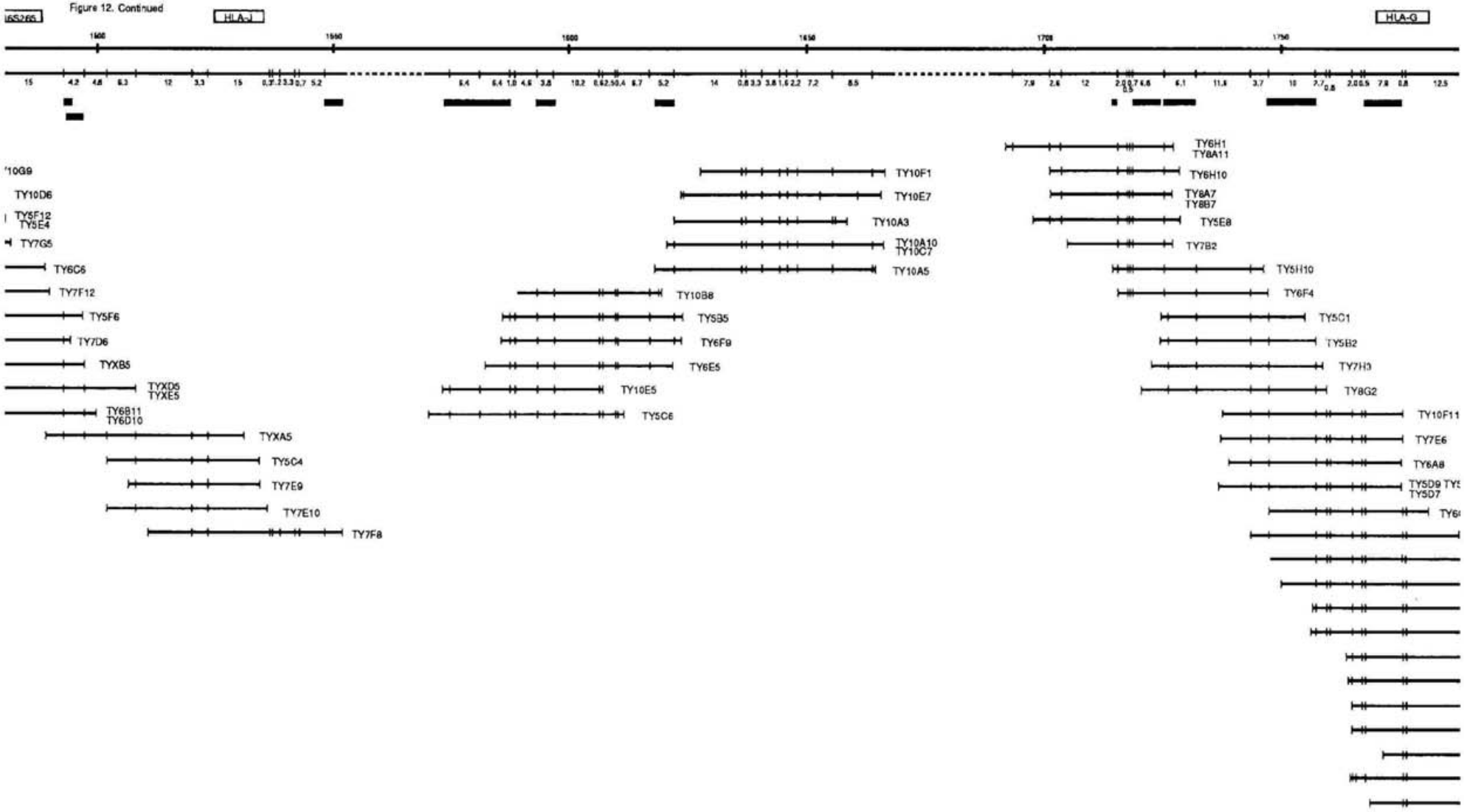
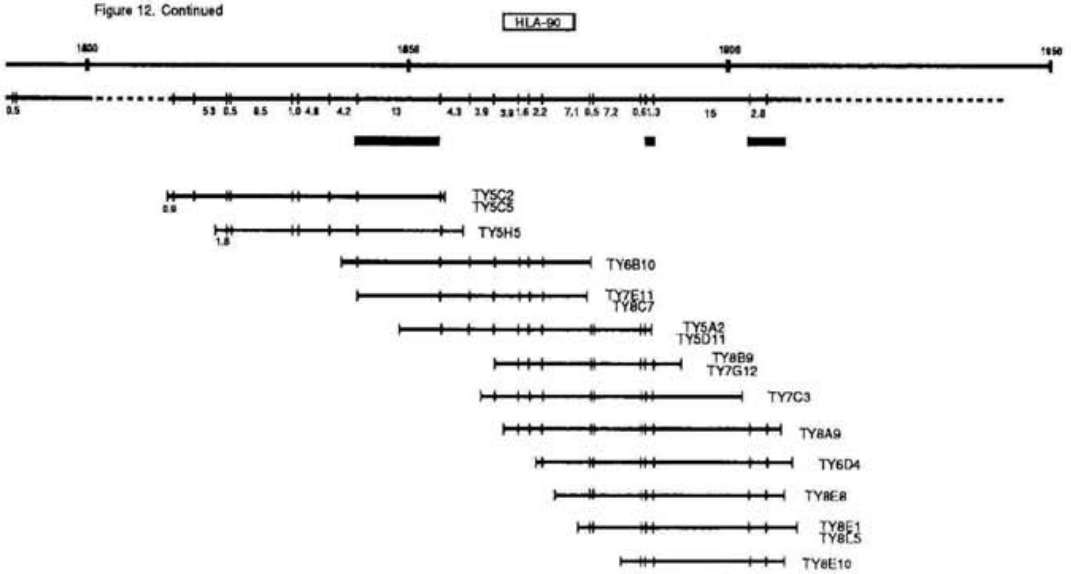


Figure 12. Continued



F4

32

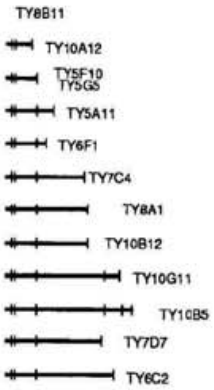


Figure 12. Continued

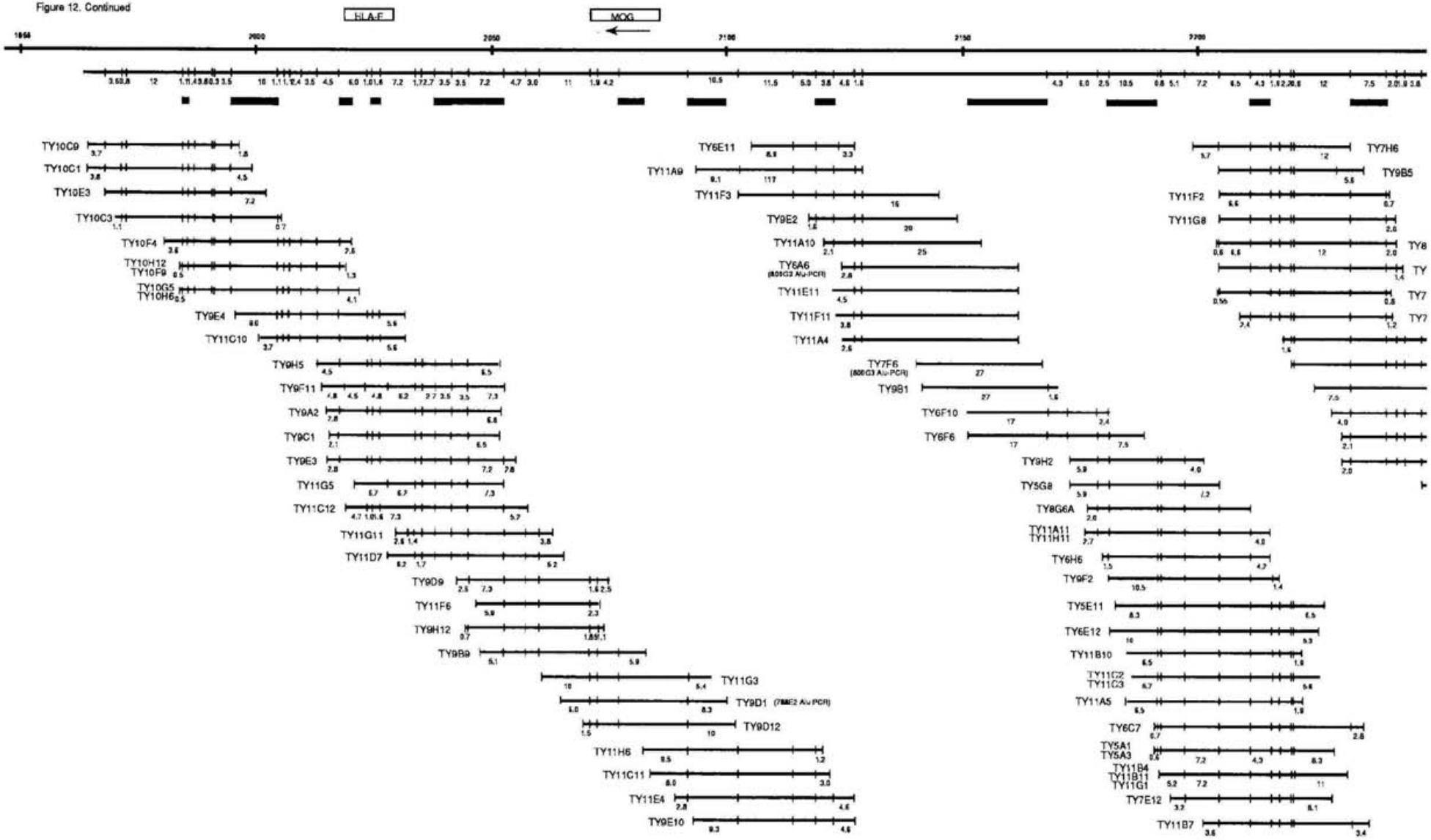
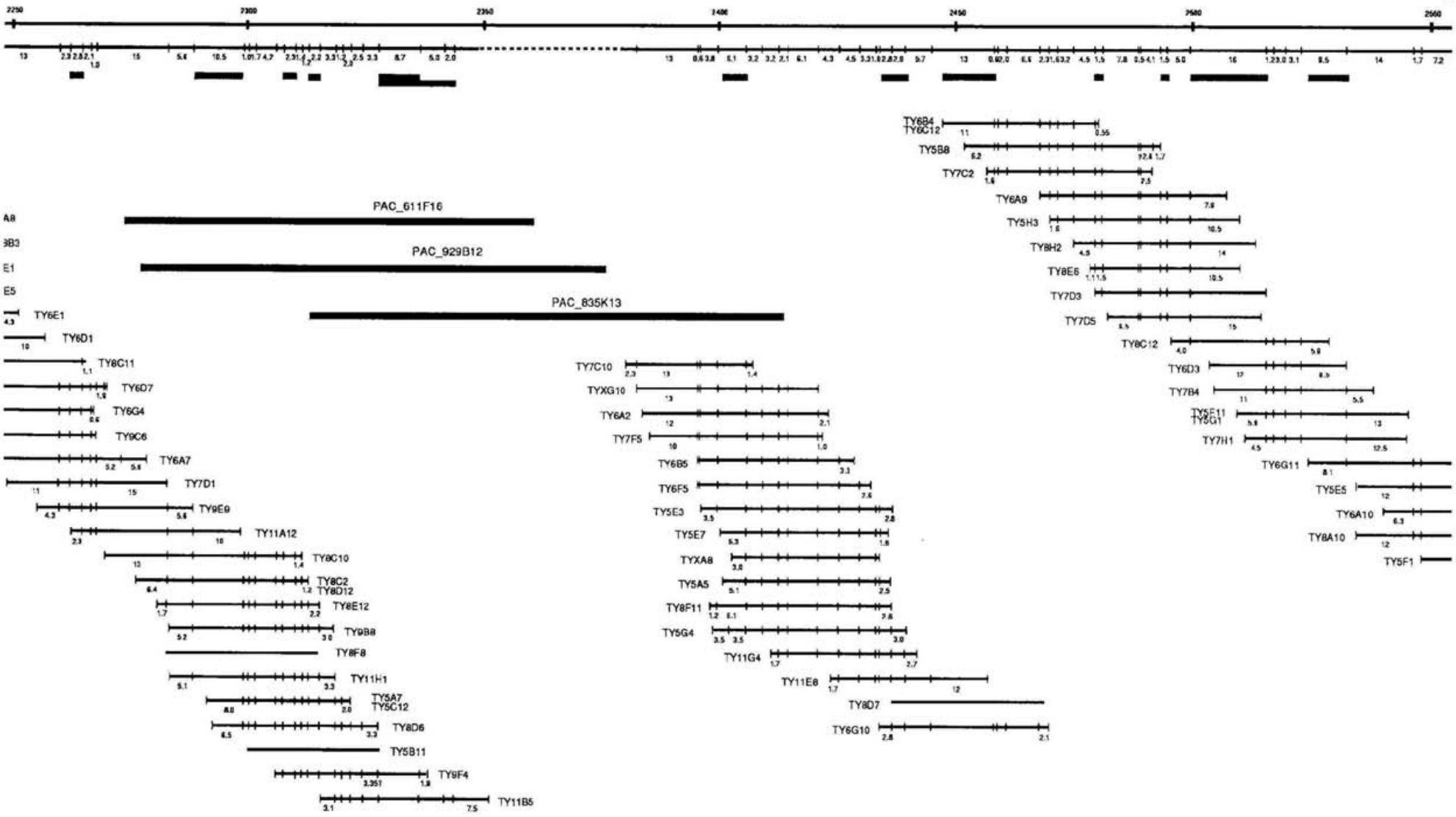


Figure 12. Continued



77

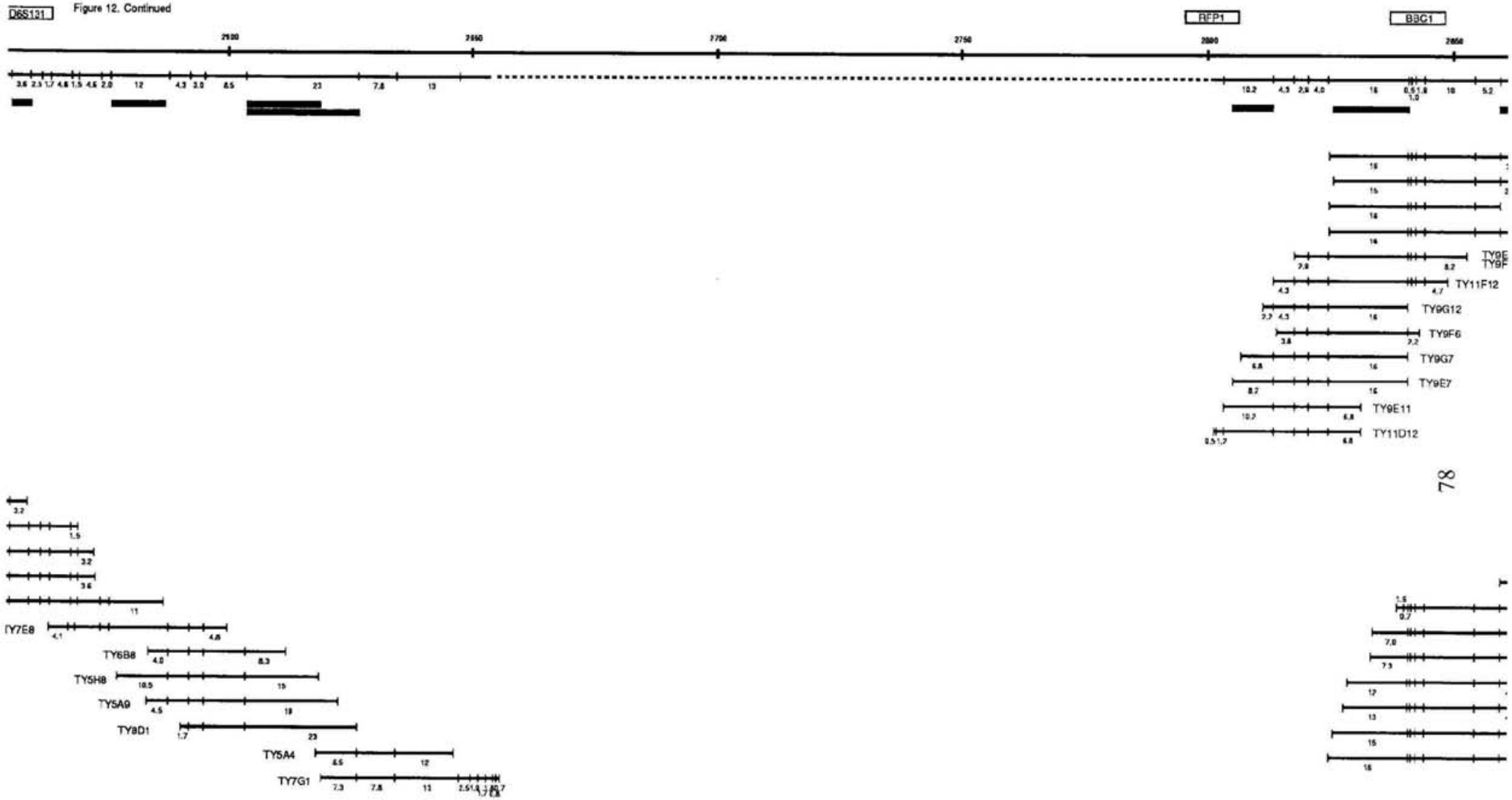

D6S131
Figure 12. Continued
RFP1
BBC1
2500
2650
2700
2750
2800
2850
TY9E
TY9F
TY11F12
TY9G12
TY9F6
TY9G7
TY9E7
TY9E11
TY11D12
TY7E8
TY6B8
TY5H8
TY5A9
TY8D1
TY5A4
TY7G1

Figure 12. Continued

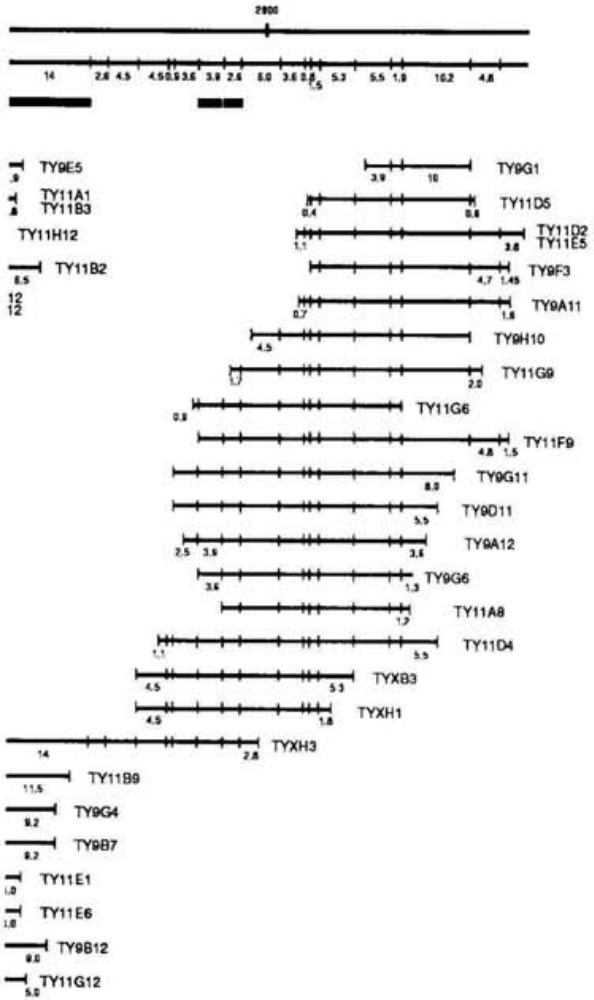
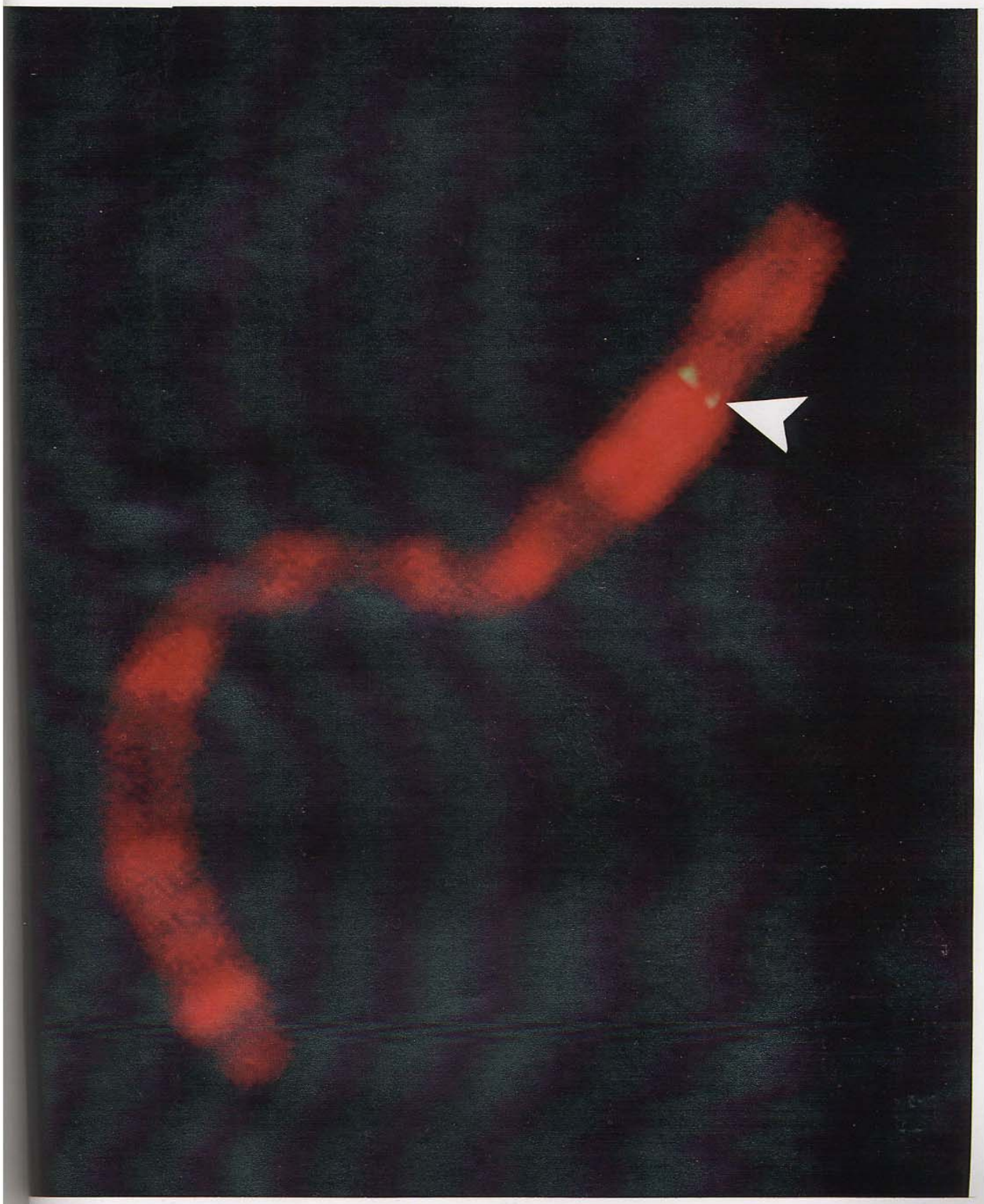


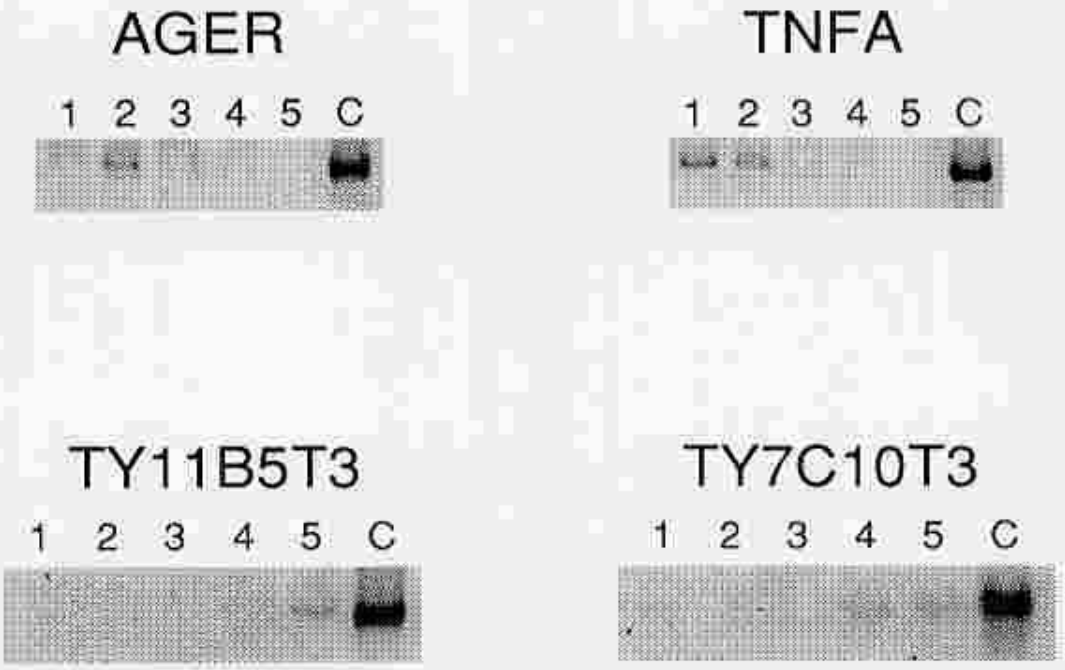
Figure 13



LEGENDS TO FIGURE 14

Figure 14. PCR products electrophoresed through an 8% polyacrylamide gel. A condition of PCR is identical to Fig. 2 except for omission of the competitor DNAs. Determination of replication timings for individual loci was not affected by the PCR conditions. The lanes represent 0-1 hour (lane 1), 1-2 hours (lane 2), 2-3 hours (lane 3), 3-4 hours (lane 4), 4-5 hours (lane 5), and control (lane 6). The control is a PCR product from the normal human genomic DNA, which is amplified using the respective primer set.

Figure 14.



LEGENDS TO FIGURES 15-19

Figure 15. Distributions of GC% and of repetitive sequences. Window size is 10 kb. Vertical lines represent repetitive sequences other than *Alu* element and LINE-1. The numbers outlined by circle indicate sequence No. around the GC% transition point listed in Table 4.

Figure 16. AG% distribution in the long human sequences containing GC% boundaries. HS49J10 and HSU91318 gave sharp high or low AG% peaks, which were located close to the GC% transition point.

Figure 17. GC% distribution in the human (A) and mouse (B) sequences contain the GC% boundary between MHC classes II and III; HSMHC3A5 and MMMHC29N9. Arrowhead indicated homologous sequence, which is found by dot-matrix plots (Fig.18).

Figure 18. Graphical dot-matrix plots of detailed comparison between human (HSMHC3A5) and mouse sequences (MMMHC29N9). The X-axis represents the mouse sequence and the Y-axis represents the human sequence. The homologous regions are indicated by the dots. The dots present at 19.9 kb, 21.7 kb, and 22.4 kb apart from the beginning of the human sequence correspond to poly A or poly T sequence in the vicinity of *Alu* element. About 1-kb region being conserved between the species is indicated with arrowheads.

Figure 19. Alignment of the 1-kb conserved nucleotide sequences found in Fig. 18. Identical nucleotides are marked by * bellow the sequence.

Figure 15.

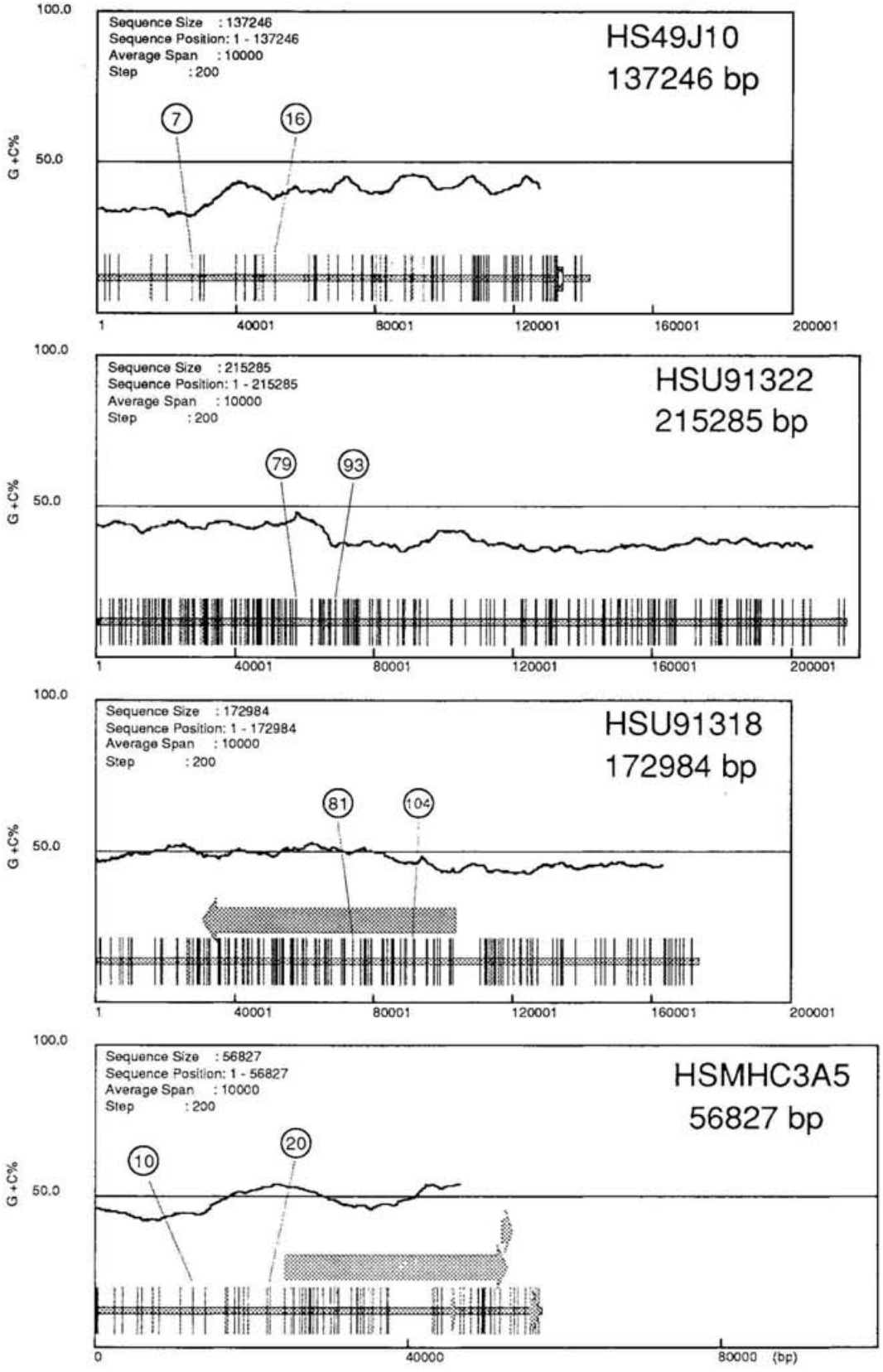
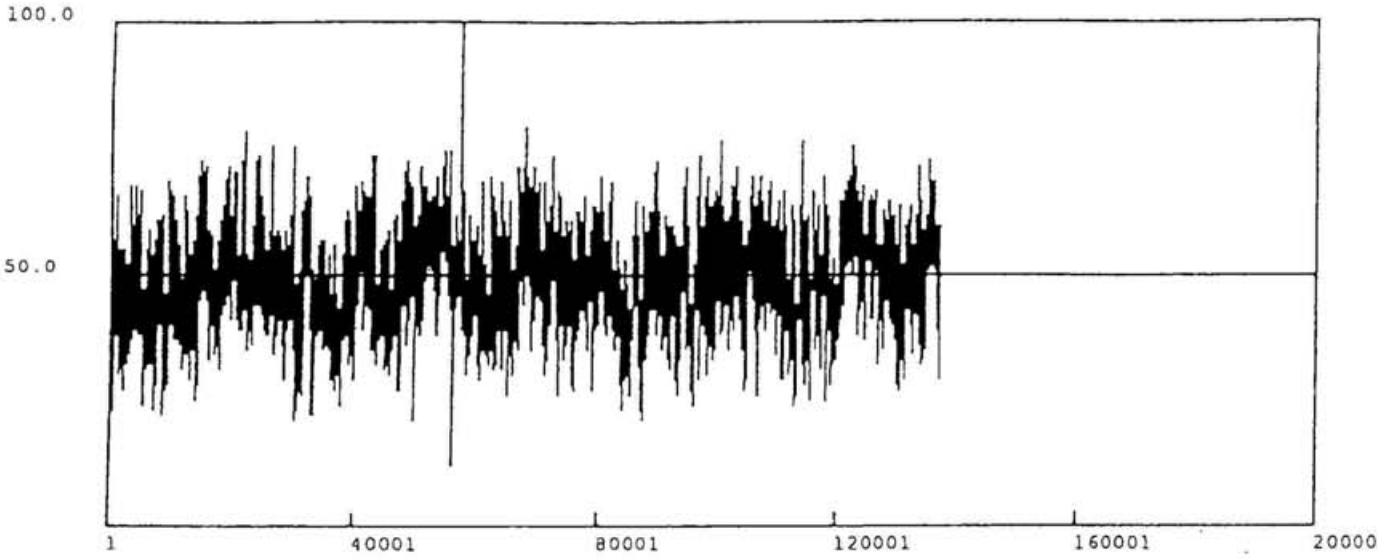

100.0
Sequence Size : 137246
Sequence Position: 1 - 137246
Average Span : 10000
Step : 200
HS49J10
137246 bp
7
16
G +C%
50.0
1
40001
80001
120001
160001
200001
100.0
Sequence Size : 215285
Sequence Position: 1 - 215285
Average Span : 10000
Step : 200
HSU91322
215285 bp
79
93
G +C%
50.0
1
40001
80001
120001
160001
200001
100.0
Sequence Size : 172984
Sequence Position: 1 - 172984
Average Span : 10000
Step : 200
HSU91318
172984 bp
81
104
G +C%
50.0
1
40001
80001
120001
160001
200001
100.0
Sequence Size : 56827
Sequence Position: 1 - 56827
Average Span : 10000
Step : 200
HSMHC3A5
56827 bp
10
20
G +C%
50.0
0
40000
80000 (bp)

Figure 16.

A

Filename : HS49J10
Sequence Size : 137246
Sequence Position: 1 - 137246
Average Span : 100
Step : 1



B

Filename : HSU91318
Sequence Size : 172984
Sequence Position: 1 - 172984
Average Span : 100

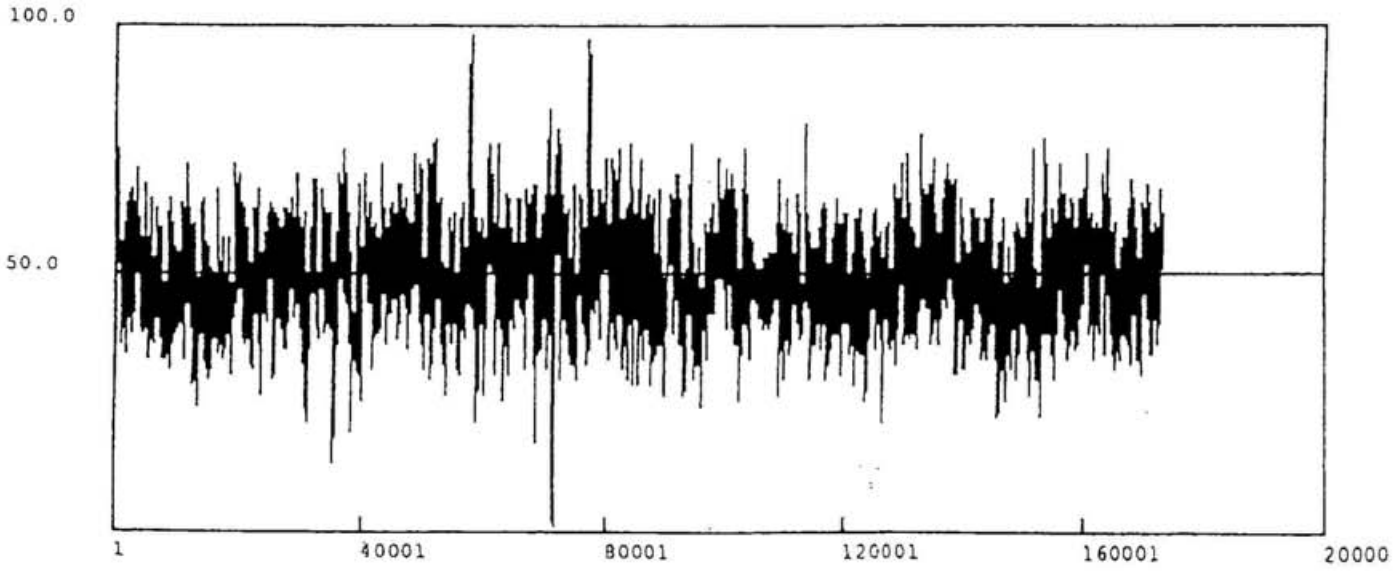
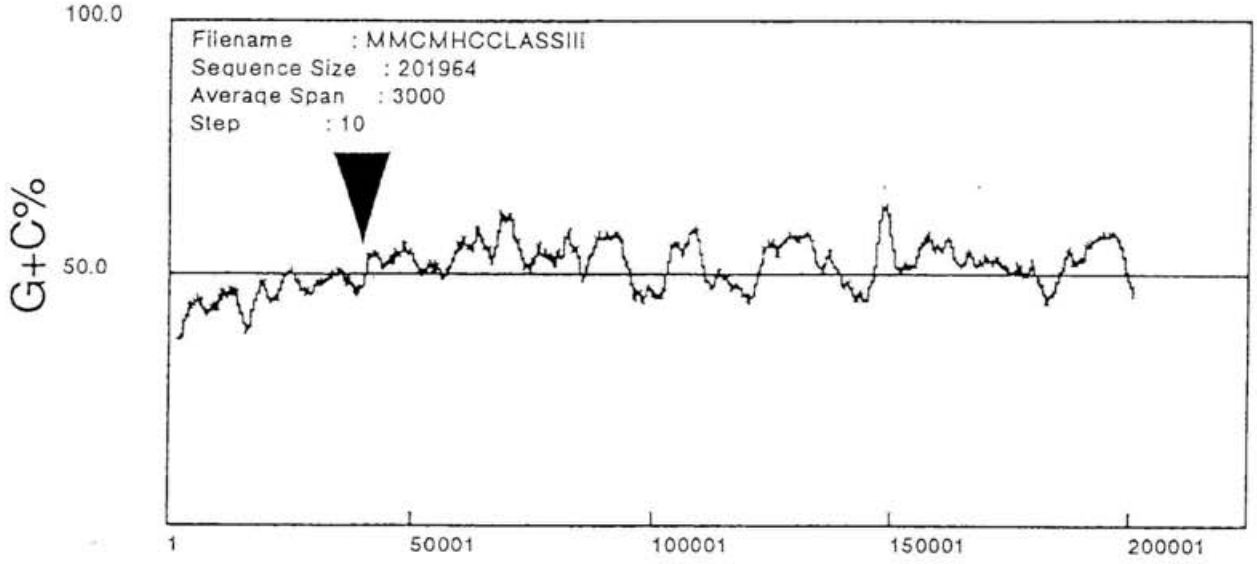


Figure 17.

A


100.0
50.0
G+C%
Filename : MMCMHCCLASSIII
Sequence Size : 201964
Average Span : 3000
Step : 10
1
50001
100001
150001
200001

B

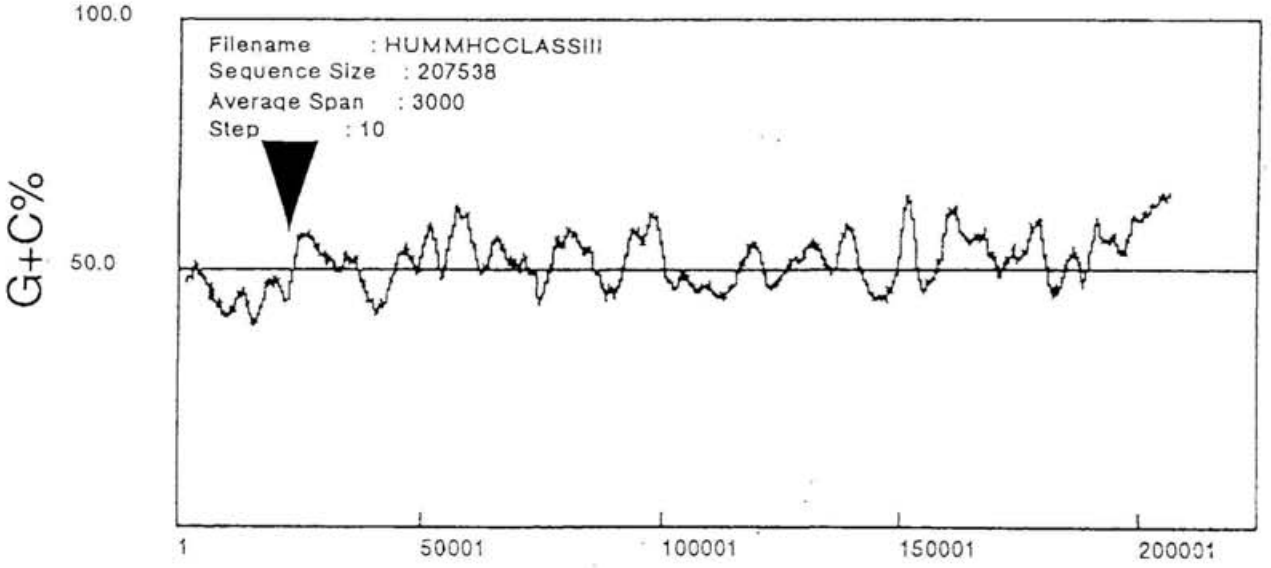

100.0
50.0
G+C%
Filename : HUMMHCCLASSIII
Sequence Size : 207538
Average Span : 3000
Step : 10
1
50001
100001
150001
200001

Figure 18.

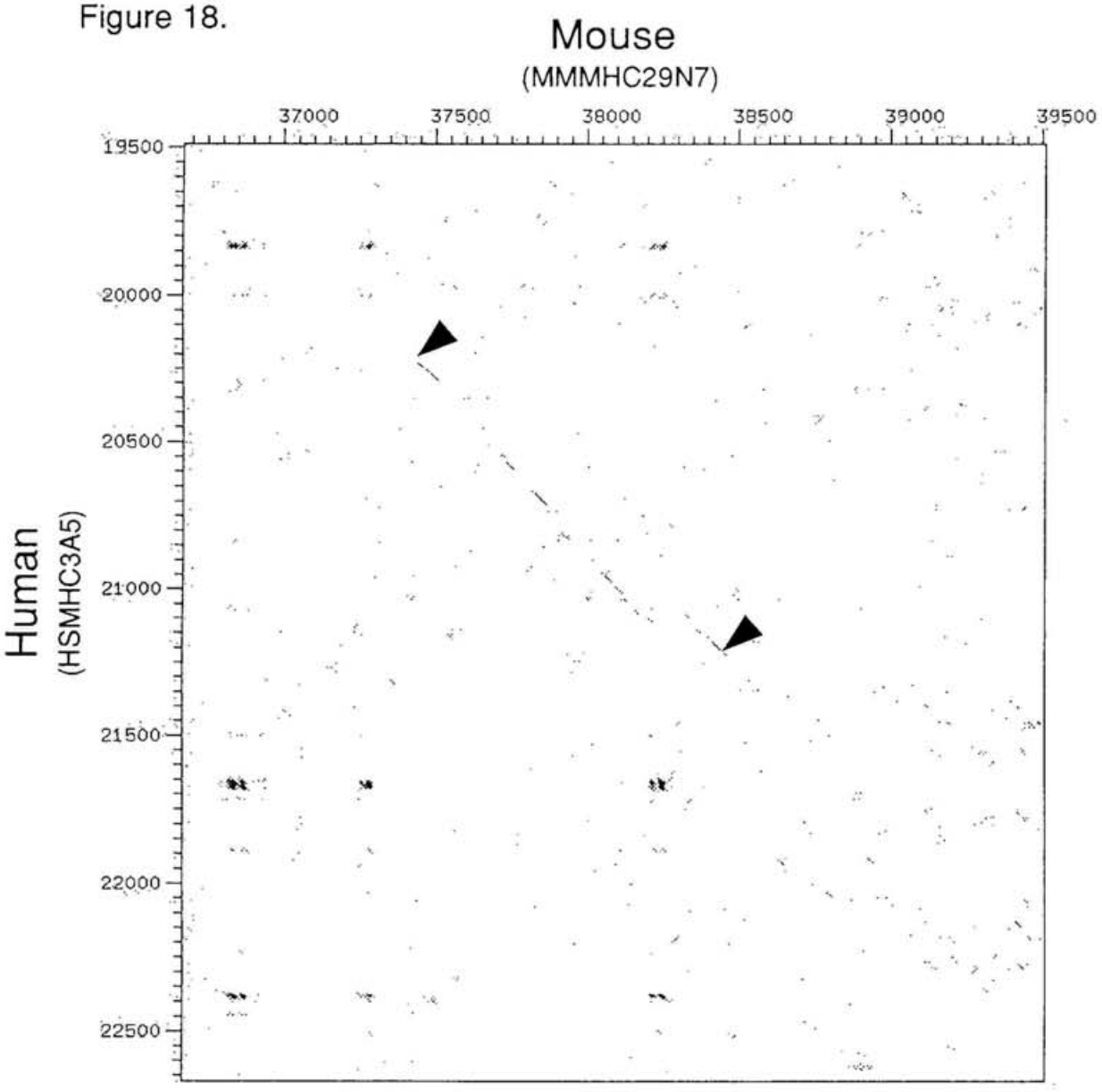
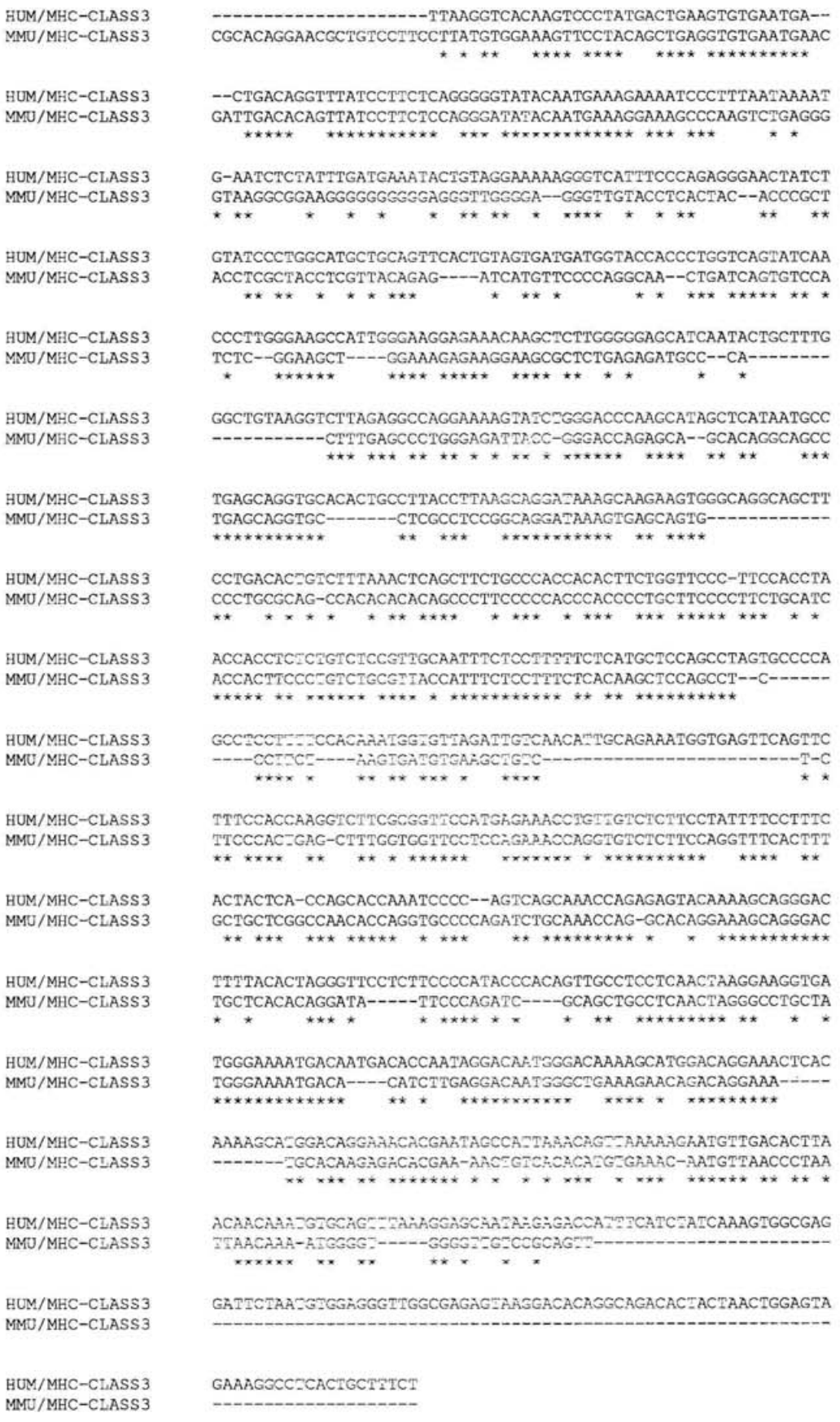

Mouse
(MMMHC29N7)
37000
37500
38000
38500
39000
39500
19500
20000
20500
21000
21500
22000
22500
Human
(HSMHC3A5)

Figure 19.


HUM/MHC-CLASS3 ----------TTAAGGTCACAAGTCCCTATGACTGAAGTGTGAATGA--
MMU/MHC-CLASS3 CGCACAGGAACGCTGTCCTTCCTTATGTGGAAAGTTCCTACAGCTGAGGTGTGAATGAAC
HUM/MHC-CLASS3 --CTGACAGGTTTATCCTTCTCAGGGGGTATACAATGAAAGAAAATCCCTTTAATAAAAT
MMU/MHC-CLASS3 GATTGACACAGTTATCCTTCTCCAGGGATATACAATGAAAGGAAAGCCCAAGTCTGAGGG
HUM/MHC-CLASS3 G-AATCTCTATTTGATGAAATACTGTAGGAAAAAGGGTCATTTCCCAGAGGGAACTATCT
MMU/MHC-CLASS3 GTAAGGCGGAAGGGGGGGGGGAGGGTTGGGGA--GGGTTGTACCTCACTAC--ACCCGCT
HUM/MHC-CLASS3 GTATCCCTGGCATGCTGCAGTTCACTGTAGTGATGATGGTACCACCCTGGTCAGTATCAA
MMU/MHC-CLASS3 ACCTCGCTACCTCGTTACAGAG----ATCATGTTCCCCAGGCAA--CTGATCAGTGTCCA
HUM/MHC-CLASS3 CCCTTGGGAAGCCATTGGGAAGGAGAAACAAGCTCTTGGGGGAGCATCAATACTGCTTTG
MMU/MHC-CLASS3 TCTC--GGAAGCT----GGAAAGAGAAGGAAGCGCTCTGAGAGATGCC--CA--------
HUM/MHC-CLASS3 GGCTGTAAGGTCTTAGAGGCCAGGAAAAGTATCTGGGACCCAAGCATAGCTCATAATGCC
MMU/MHC-CLASS3 ----------CTTTGAGCCCTGGGAGATTACC-GGGACCAGAGCA--GCACAGGCAGCC
HUM/MHC-CLASS3 TGAGCAGGTGCACACTGCCTTACCTTAAGCAGGATAAAGCAAGAAGTGGGCAGGCAGCTT
MMU/MHC-CLASS3 TGAGCAGGTGC-------CTCGCCTCCGGCAGGATAAAGTGAGCAGTG------------
HUM/MHC-CLASS3 CCTGACACTGTCTTTAAACTCAGCTTCTGCCCACCACACTTCTGGTTCCC-TTCCACCTA
MMU/MHC-CLASS3 CCCTGCGCAG-CCACACACACAGCCCTTCCCCCACCCACCCCTGCTTCCCCTTCTGCATC
HUM/MHC-CLASS3 ACCACCTCTCTGTCTCCGTTGCAATTTCTCCTTTTTCTCATGCTCCAGCCTAGTGCCCCA
MMU/MHC-CLASS3 ACCACTTCCCTGTCTGCGTTACCATTTCTCCTTTCTCACAAGCTCCAGCCT--C------
HUM/MHC-CLASS3 GCCTCCTTTTCCACAAATGGTGTTAGATTGTCAACATTGCAGAAATGGTGAGTTCAGTTC
MMU/MHC-CLASS3 ----CCTTCT----AAGTGATGTGAAGCTGTC--------------------------T-C
HUM/MHC-CLASS3 TTTCCACCAAGGTCTTCGGGGTTCCATGAGAAACCTGTTGTCTCTTCCTATTTTCCTTTC
MMU/MHC-CLASS3 TTCCCACTGAG-CTTTGGTGGTTCCTCCAGAAACCAGGTGTCTCTTCCAGGTTTCACTTT
HUM/MHC-CLASS3 ACTACTCA-CCAGCACCAAATCCCC--AGTCAGCAAACCAGAGAGTACAAAAGCAGGGAC
MMU/MHC-CLASS3 GCTGCTCGGCCAACACCAGGTGCCCCAGATCTGCAAACCAG-GCACAGGAAAGCAGGGAC
HUM/MHC-CLASS3 TTTTACACTAGGGTTCCTCTTCCCCATACCCACAGTTGCCTCCTCAACTAAGGAAGGTGA
MMU/MHC-CLASS3 TGCTCACACAGGATA-----TTCCCAGATC----GCAGCTGCCTCAACTAGGGCCTGCTA
HUM/MHC-CLASS3 TGGGAAAATGACAATGACACCAATAGGACAATGGGACAAAAGCATGGACAGGAAACTCAC
MMU/MHC-CLASS3 TGGGAAAATGACA----CATCTTGAGGACAATGGGCTGAAAGAACAGACAGGAAA-----
HUM/MHC-CLASS3 AAAAGCATGGACAGGAAACACGAATAGCCATTAAACAGTTAAAAAGAATGTTGACACTTA
MMU/MHC-CLASS3 -------TGCACAAGAGACACGAA-AACTGTCACACACTGTGAAAC-AATGTTAACCCTAA
HUM/MHC-CLASS3 ACAACAAATGTGCAGTTTAAAGGAGCAATAAGAGACCATTTCATCTATCAAAGTGGCGAG
MMU/MHC-CLASS3 TTAACAAA-ATGGGGT-----GGGGTTGTCCGCAGTT------------------------
HUM/MHC-CLASS3 GATTCTAACTGTGGAGGGTTGGCGAGAGTAAGGACACAGGCAGACACTACTAACTGGAGTA
MMU/MHC-CLASS3 --
HUM/MHC-CLASS3 GAAAGGCCTCACTGCTTTCT
MMU/MHC-CLASS3 --------------------

LEGENDS TO FIGURES 20-22

Figure 20. Physical map of human GABAB receptor gene locus in the human MHC class I. Horizontal bars represent cosmid clones mapped to this region and their names are shown on the left side. Position of restriction sites and fragment sizes for *Eco*RI and *Bam*HI are indicated in the map at the bottom of the figure. Arrows indicate genes and their transcriptional direction.

Figure 21. CLUSTAL-W alignment of the nucleotide sequence encoding the human GABAB receptor with R1a cDNA of the rat GABAB receptor (GenBank Accession No. Y10369). The nucleotide identity is 91% in the total sequence.

Figure 22. Genomic structure of the human GABAB receptor gene. Translated exons are represented by closed and hatched boxes on the line. Repetitive sequences are shown by open (the normal direction of the reported consensus sequence) and crosshatched (its complementary direction) boxes. Restriction enzyme sites are designated as: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII. Translation start (ATG) and termination (TGA) codons are shown, together with the polyadenylation signal (AATAAA).

Figure 20.

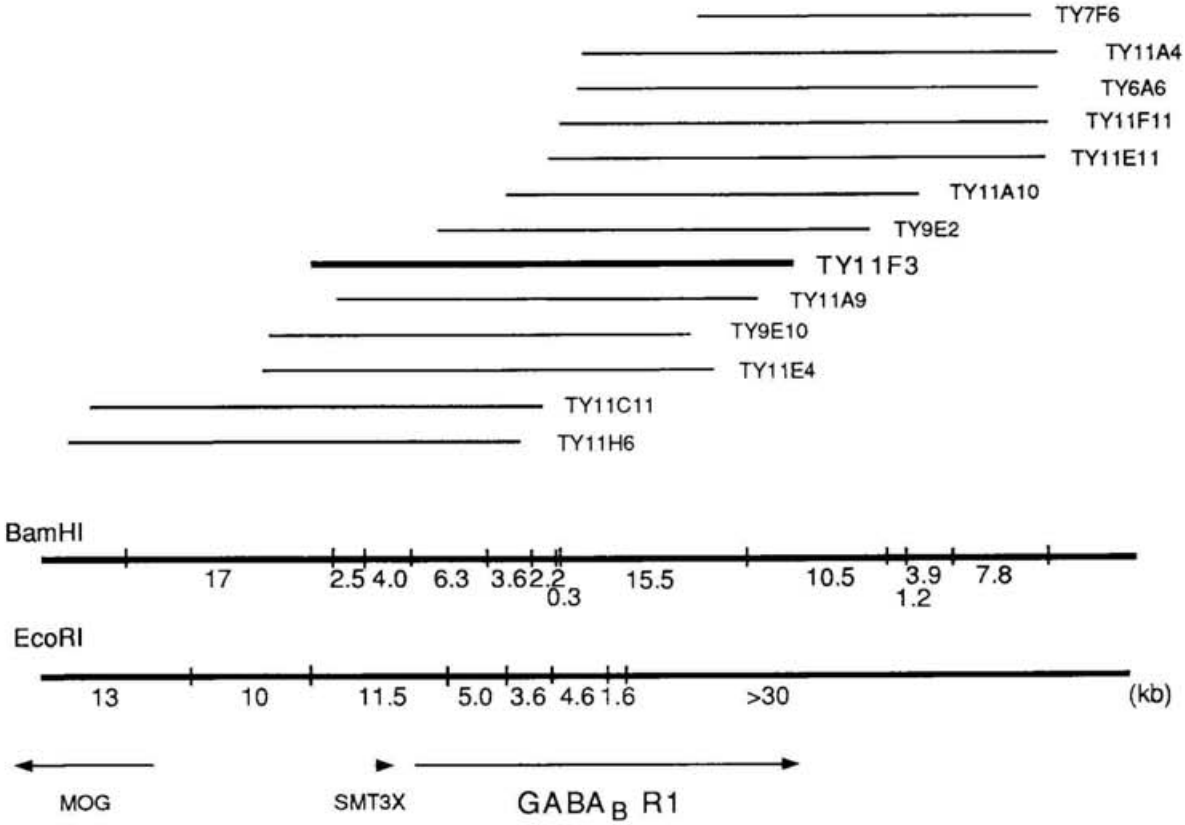
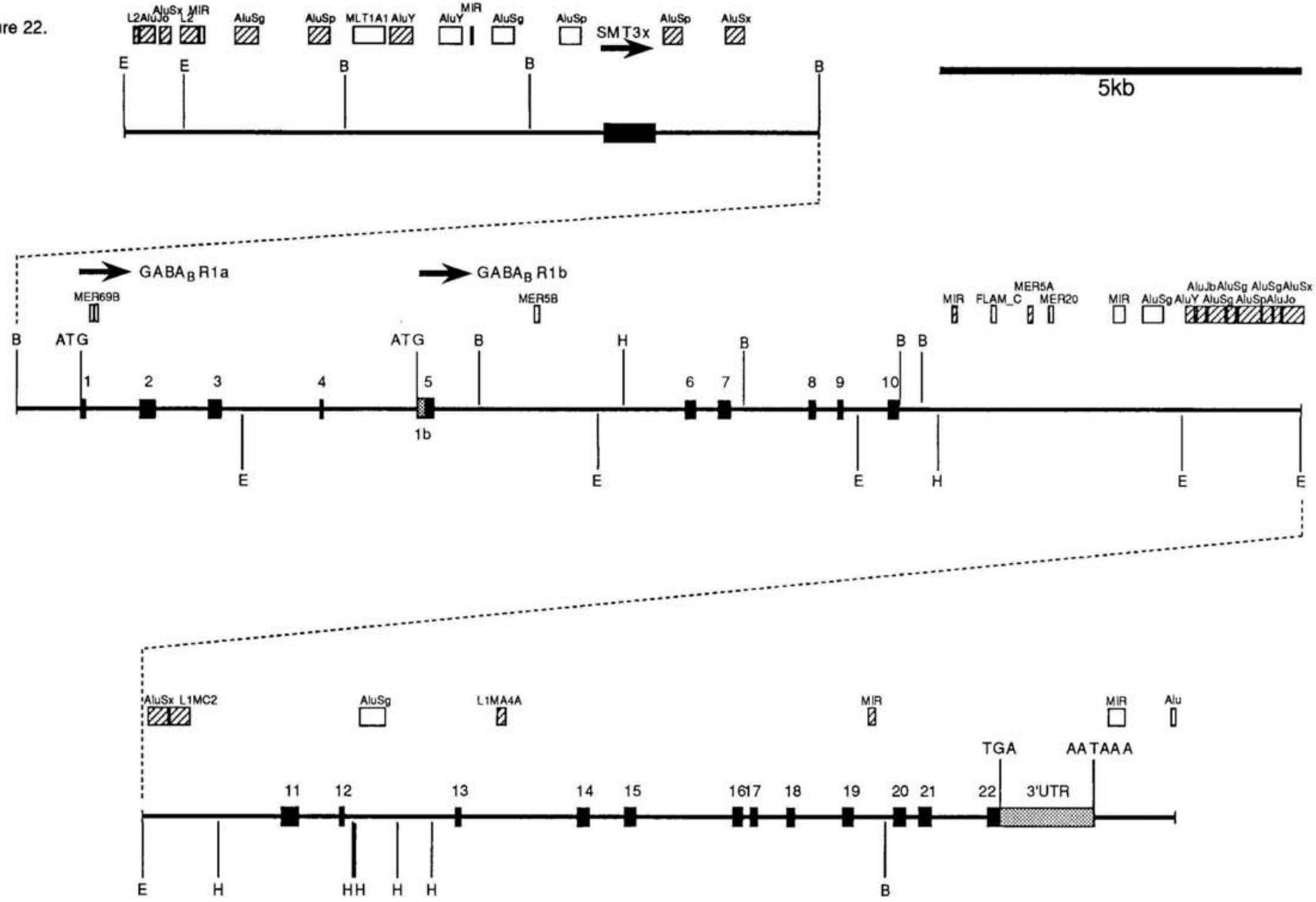

TY7F6
TY11A4
TY6A6
TY11F11
TY11E11
TY11A10
TY9E2
TY11F3
TY11A9
TY9E10
TY11E4
TY11C11
TY11H6
BamHI
17
2.5
4.0
6.3
3.6
2.2
0.3
15.5
10.5
3.9
1.2
7.8
EcoRI
13
10
11.5
5.0
3.6
4.6
1.6
>30
(kb)
MOG
SMT3X
GABA B R1

Figure 21.

Figure 22.


L2
AluJo
AluSx
L2
MIR
AluSg
AluSp
MLT1A1
AluY
AluY
MIR
AluSg
AluSp
SMT3x
AluSp
AluSx
E
E
B
B
B
5kb
GABA$_B$R1a
GABA$_B$R1b
MER69B
MER5B
MIR
FLAM_C
MER5A
MER20
MIR
AluSg
AluY
AluJb
AluSg
AluSg
AluSp
AluSg
AluSx
AluJo
B
ATG
ATG
B
H
B
B
B
1
2
3
4
5
6
7
8
9
10
1b
E
E
E
H
E
E
AluSx
L1MC2
AluSg
L1MA4A
MIR
MIR
Alu
TGA
AATAAA
11
12
13
14
15
16
17
18
19
20
21
22
3'UTR
E
H
H
H
H
B

LEGENDS TO FIGURES 23-25

Figure 23. Alignment of the amino acid sequences of the human GABAB receptor with the rat GABAB receptor. Identical residues are indicated by asterisk and similar residues are indicated by dot. The amino acid identity for R1a (A) and R1b (B) isoforms is 98% and 97%, respectively. Inverted triangles indicate the 5'-end of exon 5. Both isoforms are identical between the two from exon 5 to the end.

Figure 24. Alignment of transmembrane domains of 7TM G-protein coupled receptors. Boxes represent putative transmembrane domains (TM1-7). Asterisks and dots indicate identical and similar residues among the sequences. Inverted triangles indicate cysteine residues that probably form a disulfide bond.

Figure 25. Structure of the human GABAB receptor isoforms. A) Translated exons are designated as closed (common between the isoforms) and hatched (isoform-specific) boxes. The nucleotide identity (%) for each exon between rat and human sequences is listed above the box. The number under the boxes represents exon number. B) Restriction map of cosmid TY11F3 . C) Transcription of the R1a isoform initiates at the first promoter and the R1b isoform at the second promoter.

Figure 23.


A
hsgababla MLLLLLLAPLFLRPPGAGGAQTPNATSEGCQIIHPPWEGGIRYRGLTRDQVKAINFLPVD
rngababla *-*****V******L**********
hsgababla YEIEYVCRGEREVVGPKVRKCLANGSWTDMDTPSRCVRICSKSYLTLENGKVFLTGGDLP
rngababla **********
hsgababla AIDGARVDFRCDPDFHLVGSSRSICSQGQWSTPKPHCQVNRTPHSERRAVYIGALFPMSG
rngababla *******E***********V**********
hsgababla GWPGGQACQPAVEMALEDVNSRRDILPDYELKFIHHDSKCDPGQATKYLYELLYNDPIKI
rngababla **********L**********
hsgababla ILMPGCSSVSTLVAEAARMWNLIVLSYGSSSPALSNRQRFPTFFRTHPSATLHNPTRVKL
rngababla **********
hsgababla FEKWGWKKIATIQQTTEVFTSTLDDLEERVKEAGIEITFRQSFFSDPAVPVKNLKRQDAR
rngababla **********
hsgababla IIVGLFYETEARKVFCEVYKERLFGKKVWFLIGWYADNWFKIYDPSINCTVDEMTEAVE
rngababla **********T*********E*******
hsgababla GHITTEIVMLNPANTRSISNMTSQEFVEKLTKRLKRHPEETGGFQEAPLAYDAIWALALA
rngababla **********
hsgababla LNKTSGGSGRSGVRLEDFNYNNQTITDQIYRAMNSSSFEGVSGHVVFDASGSRMAWTLIE
rngababla *******G**********
hsgababla QLQGGSYKKIGYYDSTKDDLSWSKTDKWIGGSPPADQTLVIKTFRFLSQKLFISVSVLSS
rngababla **********
hsgababla LGIVLAVVCLSFNIYNSHVRYIQNSQPNLNNLTAVGCSLALAAVFPLGLDGYHIGRNQFP
rngababla **********S***
hsgababla FVCQARLWLLGLGFSLGYGSMFTKIWWVHTVFTKKEEKKEWRKTLEPWKLYATVGLLVGM
rngababla **********
hsgababla DVLTLAIWQIVDLLHRTIETFAKEEPKEDTDVSILPQLEHCSSRKMNTWLGIFYGYKGLL
rngababla **********P**********K**********
hsgababla LLLGIFLAYETKSVSTEKINDHRAVGMAIYNVAVLCLITAPVTMILSSQQDAAFAFASLA
rngababla **********
hsgababla IVFSSYITLVVLFVPKMRRLITRGEWQSEAQDTMKTGSSTNNNEEEKSRLLEKENRELEK
rngababla **********T**********
hsgababla IIAEKEERVSELRHQLQSRQQLRSRRHPPTPPEPSGGLPRGPPEPPDRLSCDGSRVHLLYK
rngababla **********D*********S**********
B
hsgababl1b MGPGAPFARVGWPLPLLVVMAAGVAPVWASHSPHLPRPHSRVPPHPSSK
rngababl1b ****G*CTP********L**********P*****P********P

Figure 24.

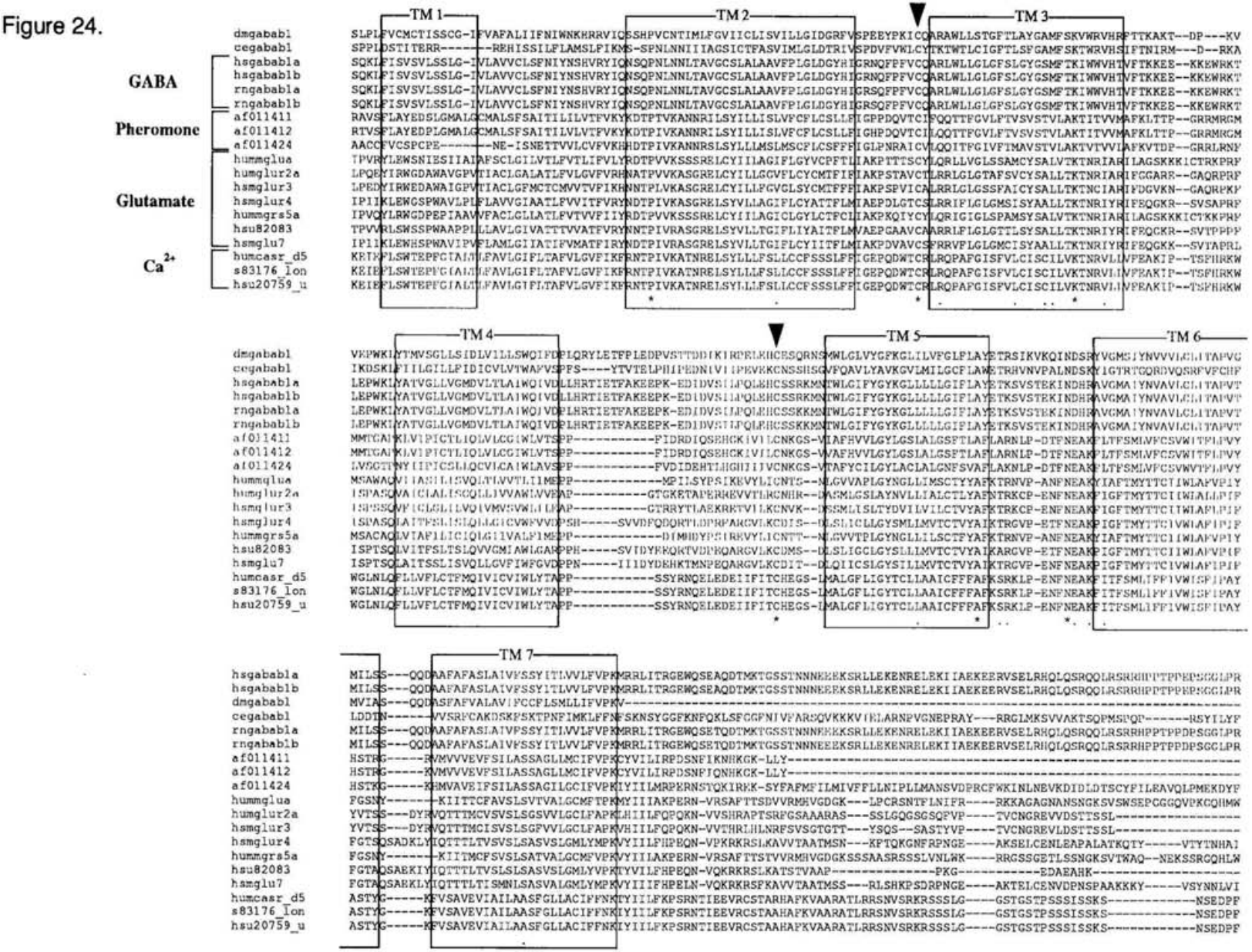
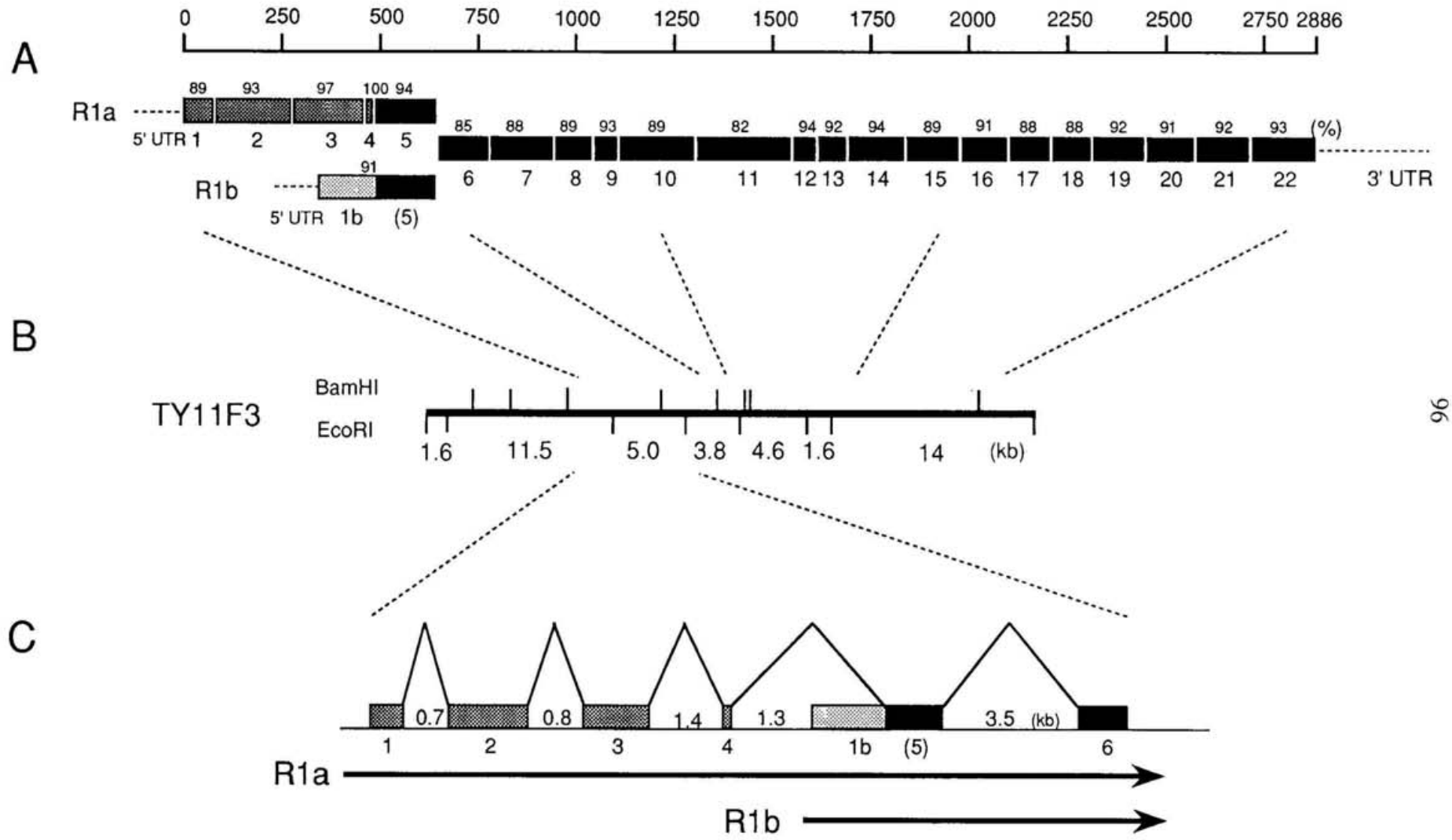

GABA
Pheromone
Glutamate
Ca2+
TM 1
TM 2
TM 3
TM 4
TM 5
TM 6
TM 7
dmgabab1
cegabab1
hsgabab1a
hsgabab1b
rngabab1a
rngabab1b
af011411
af011412
af011424
hummglua
humglur2a
hsmglur3
hsmglur4
hummgrs5a
hsu82083
hsmglu7
humcasr_d5
s83176_lon
hsu20759_u

Figure 25.


A
0 250 500 750 1000 1250 1500 1750 2000 2250 2500 2750 2886
R1a
89 93 97 100 94
5' UTR 1 2 3 4 5
85 88 89 93 89 82 94 92 94 89 91 88 88 92 91 92 93 (%)
6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
3' UTR
R1b
91
5' UTR 1b (5)
B
TY11F3
BamHI
EcoRI
1.6 11.5 5.0 3.8 4.6 1.6 14 (kb)
C
0.7 0.8 1.4 1.3 3.5 (kb)
1 2 3 4 1b (5) 6
R1a
R1b

LEGENDS TO FIGURES 26-28

Figure 26. The promoter regions of the human GABAB gene. The nucleotide sequences of the first (A) and second (B) promoter. The predicted promoter sequence is represent by shade and the transcription start site is indicated with inverted triangles, and the first exon is indicated in bold. Low-complexity sequences are in small letter with underline and consensus motifs for AP2, Sp1, CRE, and GCF are indicated by the boxes with small letters.

Figure 27. Characteristics of the promoter region. Position of AP2, CRE, GCF, Sp1 motifs are shown with dots (upper panel). The G+C% profile is plotted with a window width of 200 nt (lower panel). Two CpG islands (GC% > 70%; Obs/Exp ratio > 0.6) are identified.

Figure 28. A model of alternative promoter usage whereby the GABA signal affects gene expression for GABAB receptor via CRE (cAMP response element). Since GABAB receptors negatively couple to adenylate cyclase, stimulation of the receptor with GABA reduces the activity of CREB whose activation is regulated by cAMP level. Therefore, expression from the second promoter with CRE, is thought to be autoregulated (A). In this model, expression of the R1a isoform is interfered by CREB and/or other unknown factors which regulate the transcription from the second promoter (B, upper). When the activity of CRE-bind protein decreases, expression of R1a relatively increases (B, lower). We propose a model that the activation of GABAB receptor differentiates the number and ratio of two isoforms via CRE- and CRE-mediated regulatory system possibly modulating synaptic transmission (C).

Figure 26.

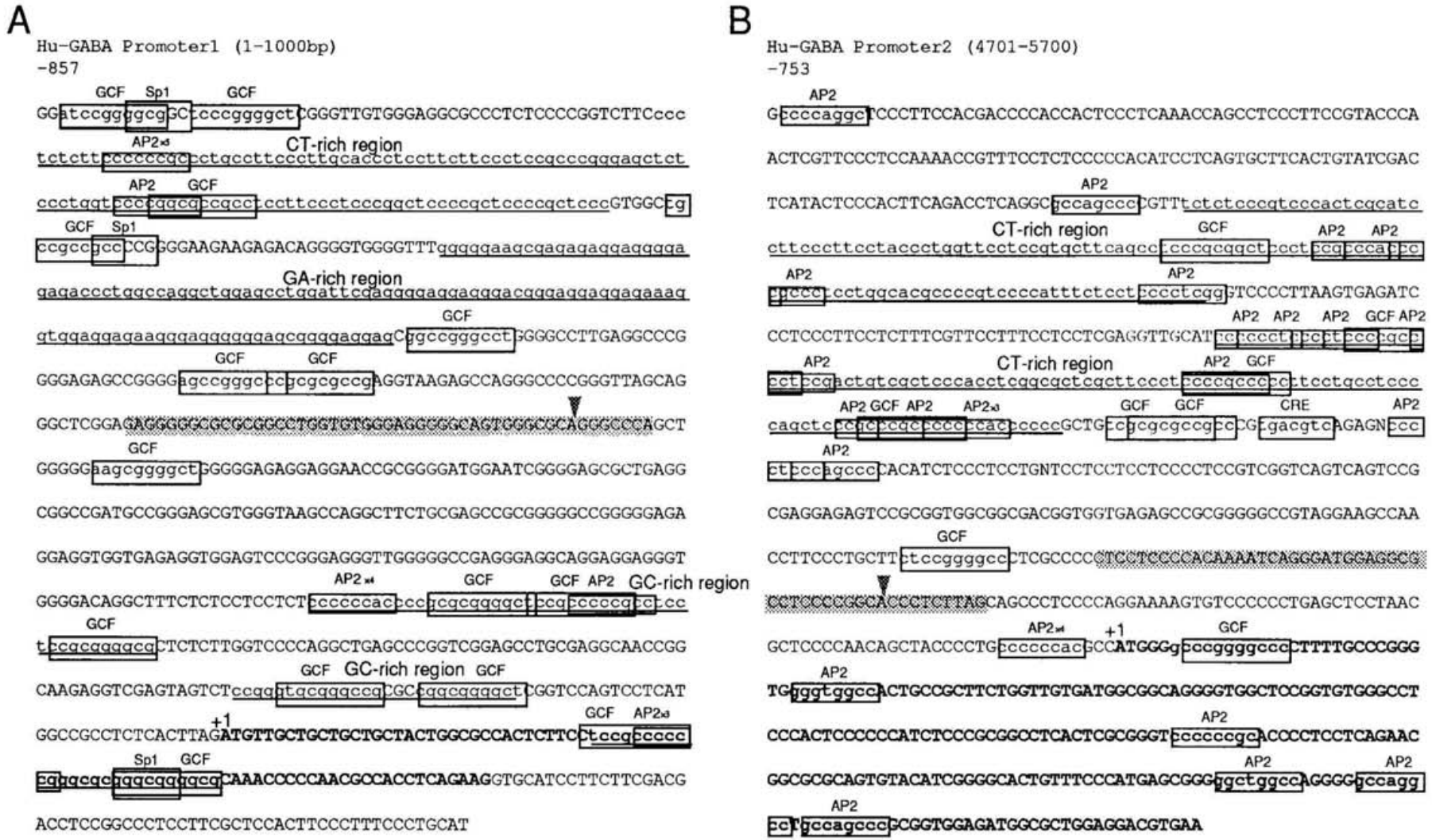

A
Hu-GABA Promoter1 (1-1000bp)
-857
GCF
Sp1
GCF
AP2
CT-rich region
AP2
GCF
GCF
Sp1
GA-rich region
GCF
GCF
GCF
GCF
AP2
GC-rich region
GCF
GCF
GC-rich region
GCF
+1
GCF
AP2
Sp1
GCF
B
Hu-GABA Promoter2 (4701-5700)
-753
AP2
CT-rich region
GCF
AP2
AP2
AP2
AP2
AP2
AP2
AP2
CT-rich region
AP2
GCF
AP2
GCF
AP2
AP2
GCF
GCF
CRE
AP2
AP2
GCF
+1
GCF
AP2
AP2
AP2
AP2

Figure 27.

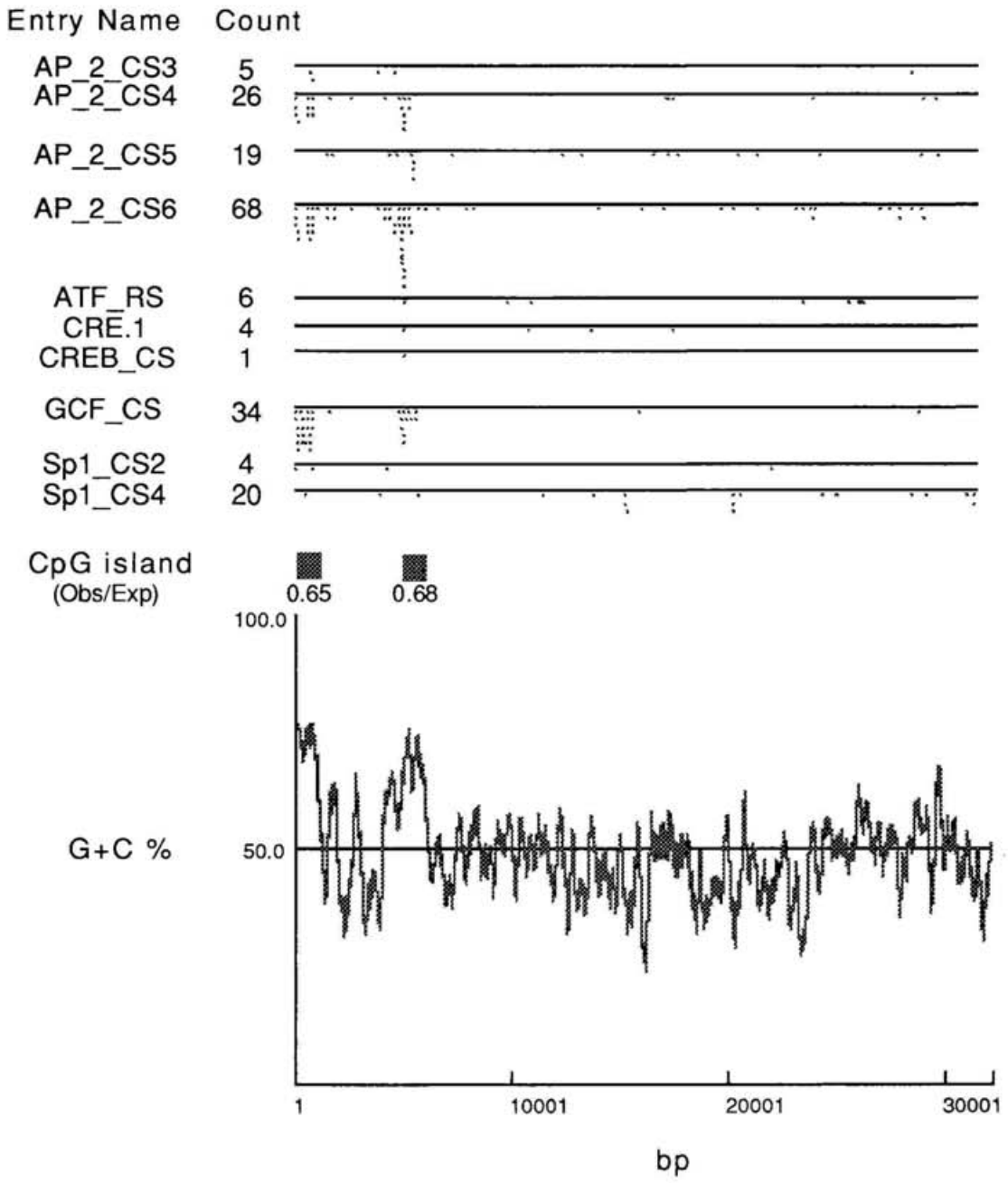

Entry Name
Count
AP_2_CS3
5
AP_2_CS4
26
AP_2_CS5
19
AP_2_CS6
68
ATF_RS
6
CRE.1
4
CREB_CS
1
GCF_CS
34
Sp1_CS2
4
Sp1_CS4
20
CpG island
(Obs/Exp)
0.65
0.68
100.0
G+C %
50.0
1
10001
20001
30001
bp

Figure 28.

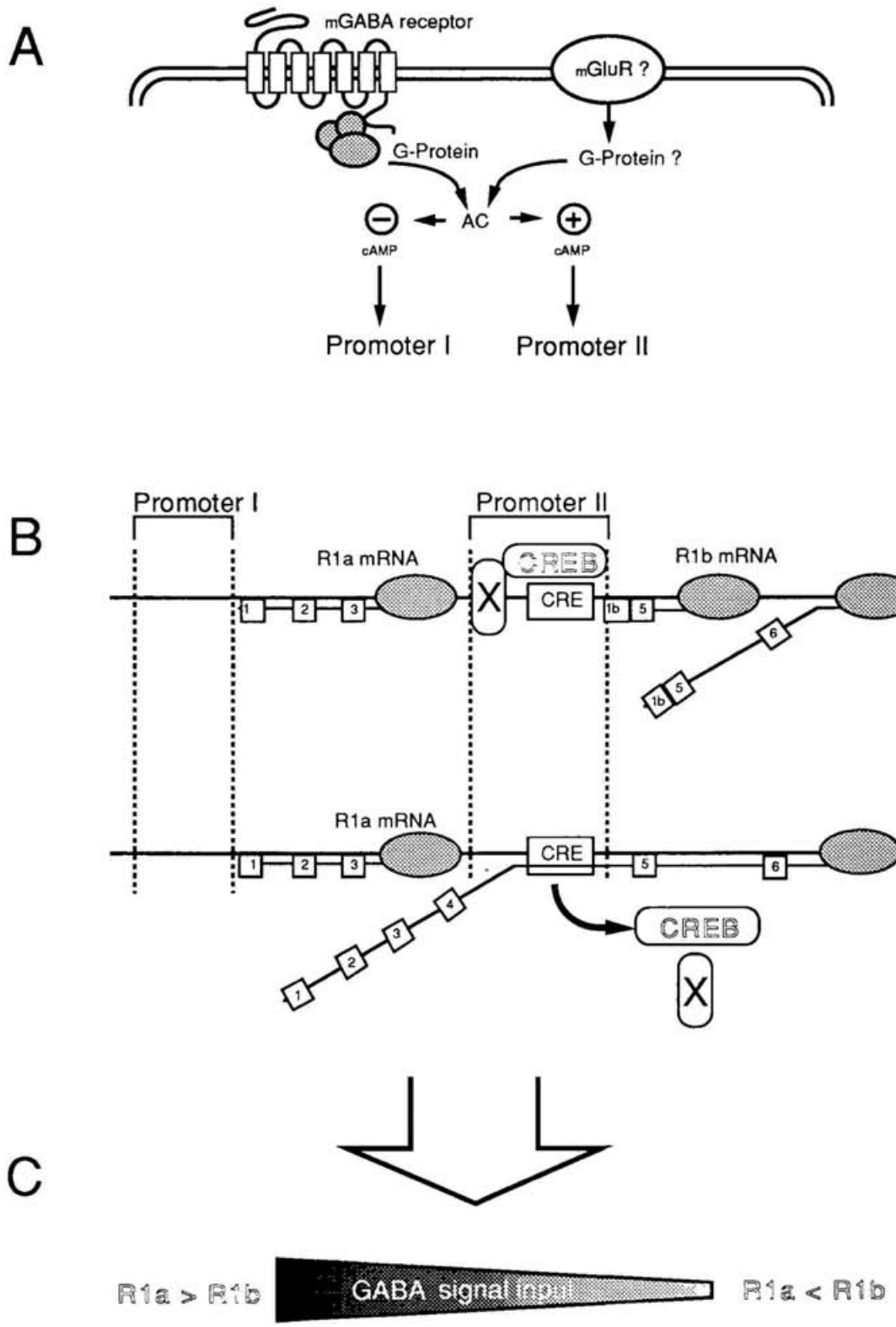

A
mGABA receptor
mGluR ?
G-Protein
G-Protein ?
AC
cAMP
cAMP
Promoter I
Promoter II
B
Promoter I
Promoter II
R1a mRNA
CREB
CRE
R1b mRNA
1
2
3
X
1b
5
6
R1a mRNA
CRE
CREB
C
R1a > R1b
GABA signal input
R1a < R1b

Table 1. Similarity of genes on chromosomes 6, 9, 1, and 19 (Sugaya *et al.*, 1997)

Chr. 6		Chr. 9		Chr. 1		Chr. 19	
RXRB	6p21.3	RXRA	9p34	RXRG	1q22-q23		
COL11A2	6p21.3	COL5A1	9p34.2-34.3	COL11A1	1p21		
TAP1, TAP2	6p21.3	ABC2	9p34				
LMP2, LMP7	6p21.3	PSMB7	9p34.11-34.12				
NOTCH4	6p21.3	NOTCH1	3p34.3	NOTCH2	1p13-p11	NOTCH3	19p13.2-p13.1
PBX2	6p21.3	PBX3	3q33-34	PBX1	1q23		
TNX	6p21.3	HXB	3q32-34	TNR	1q25-q31		
CYP21	6p21.3					CYP2 family[a]	19q13.2
C2, C4, C4B	6p21.3	C5	9q33			C3	19q13.2
HSPA1, HSPA1L	6p21.3	GRP78	9q33-34.1	HSPA6, HSPA7	1cen-qter, 1cen-qter		
VARS2	6p21.3	VARS1	9				
POU5F1	6p21.3			POU2F1	1q22-q23	POU2F	19
		CACNL1A5	9q34	CACNL1A6	1q25-q31, 1q31-q32	CACNL1A4	19p13
				CACNL1A3			
		ABL	9q34.1	ABL2	1q24-25		
		AK1, AK3	9q34.1, 9pter-p13	AK2	1q34		
		PTGS1	9q32-q33.3	PTGS2	1q25.2-q25.3		
		TAL2	9q31-q32	TAL1	1p32		
		TPM2	9p13	TPM3	1q22-q23		
TCP11	6p21.3-p21.2			TRIC5	1q23		
		VAV2	9q34			VAV1	19p13.3
		JAK2	9p24	JAK1	1p32.3-p31.3		
				TNN11	1q32	TNN13	19p13.2-q13.2
				TNNT2	1q3	TNNT1	19q13.4

All mapping information was obtained from the Genome Database.

[a] The CYP2 family consists of CYP2A, CYP2A13, CYP2A6, CYP2A7, CYP2A7P1, CYP2A7P2, CYP2B, BYP2B6, and CYP2B7

Table 2. Primers used for PCR

locus	sequence (5'→3')
DRA	AAGAACATGTGATCATCCAG
	GCGCTTTGTCATGATTTCCA
CP	TGCACCTACTATATACCCAT
	TCTTTCACCGGGGGACCTGT
PL	CATTAGTTATTTGCTGGGAA
	AGTGCCTAGCACAGTGCACA
LINE CLUSTER	AGTGTTCCTACACTCTACTT
	TTTATTGAGACCTAGCTTAC
PN112	CTCTTCGTCCTGAATGTGTC
	AGCAGCAAGAAGTATTCCTA
PCD	CCACCTATGACGTAGCCTTG
	TTGTTGAAGAAATGGCTTGT
NA	CACAACCAGTGTCACTGGAG
	TGTCAGGTTGGGGGAGAGGC
SR	TGGGACTGTAACAAAAGAGC
	ATGGCCATCAAATACATGAA
NCT	AACCTGTGTACCCCTTATTC
	CCACACCCCTACCATCTCTA
CTG REPEATS	TTAGGAGGATGATCAGTGGG
	GTTCTTCCTGGAGGTGGGCA
INT3A	GAATCCAGACAACTGTGTCA
	AGGGGACGAGGGCTAAGGCT
INT3B	AAGACAAGGGTGGGCTGGAT
	GGACTCAGGCAGTGGGACAA
AGER	AACACAGGCCGGACAGAAGC
	GGTAGACACGGACTCGGTAG
TNX	CCCAGTATGCTCTAACCTCC
	AGGTTAATGCGGTGGGTGAA
TNFA	CCAATGCCCTCCTGGCCAAT
	GCAGAGAGGAGGTTGACCTT

Table 3. Primers used for PCR

loucs	sequence (5'→3')
TY11B5T3	TGGCAGCTTTTGCCGTGGTCTAC
	AGCAGAGAATGGTTGACTCTGATC
TY7C10T3	GGAAAAGCAGTCTATCATTGGAGGG
	GGTTCTGACATGTCTAGAAGTCAAG

Table 4. Repetitive sequence in the GC% boundaries.

HSMHC3A5 [a]		HS49J10		HSU91322		HSU91318	
No.	repeat seq.	No.[b]	repeat seq.	No.[b]	repeat seq.	No.[b]	repeat seq.
11	MER102	7	Poly T	79	MIR	81	MIR
12	taa	8	Poly T	80	MIR2	82	MIR
13	MER71	9	MIR2	81	taaa	83	MIR
14	MIR2	10	MIR	82	MIR	84	(GGA)n
15	MIR2	11	MIR	83	caa	85	(GGAA)n
16	MIR2	12	MER5A	84	MIR	86	(GGGA)n
17	MIR2	13	MIR	85	MLT2D	87	(GGA)n
18	MIR2	14	MLT1B	86	MLT2D	88	MIR
19	FRAM	15	MIR	87	MLT2CB	89	MIR
00		16	MIR	88	MIR	90	MIR2
				89	MIR	91	MLT1B
				90	cgg	92	MLT1G
				91	cggg	93	MLT1F
				92	GC-rich	94	MIR
				93	AT-rich	95	MIR
				94	MER5A	96	AT_rich
						97	MIR2
						98	MIR
						99	MIR
						100	MIR
						101	MIR
						102	MIR
						103	MIR
						104	MIR

[a] This sequence contain GC% boundary region between human MHC classes II and III

[b] The numbers corresponding to numbers in Fig. 15

Table 5. Genes in and around MHC class I region.

Clones	Gene or Sequences	Accession Number	Probability (BLAST)
TY9G11	Phe-tRNA	Z84474	5.8e-92
TY9G4	yu58h04.s1	H80869	2.0e-24
TY11D12	y173g09.r1 (RFP)	H06469	3.9e-99
TY7E8	olfactry receptor (OLF1) like	U56420	1.5e-35
TY8E6	MEA11 like	U73682	1.4e-44
TY7C10	RCK like	D17532	1.3e-62
TY9B8	olfactry receptor like	Y14442	1.2e-49
TY7E5	RASH like	J00277	1.5e-36
TY9H2	mrg=mas related gene like	M13150	0.0026
TY11A5	diubiqutin	N33920	2.5e-175
TY11A5	TRE17 like	X71370	1.9e-12
TY11A10	$GABA_B$ receptor	Y10369	1.8e-46
TY11A10	olfactory receptor (FAT11) like	Z84051	9.6e-233
TY11H6	MOG	Z48051	4.2e-134
TY11D7	P5-1	L06175	1.2e-147
TY11D7	zl41a03.s1	AA152177	1.8e-40
TY10E3	HCGIV	X81005	1.9e-23
TY5C4	HLA-J	M80469	1.8e-239
TY7D6	GT475 like	X90537	1.7e-34
TY7G5	GT2530	X90532	1.2e-144
TY7G5	zd29f12.r1	W61049	4.9e-127
TY7H7	RING protein (RFB30)	Y07828	2.4e-182
TY5D8	yu58h04.r1	H80969	5.7e-130
TY5D8	C-Obc09 like	Z45994	1.2e-22
TY9H5	yx73b09.s1	N22959	1.3e-57
TY11G3	SMT3B like	X99585	7.4e-40
TY6A10	zt05e06.r1like	AA280341	8.5e-13
TY6B10	HLA-90	M96333	5.5e-259

Table 6. Exon-Intron Organization of the GABA$_B$ receptor Gene

Exon		Sequence at exon-intron junction and intron size		
No.	(bp)	5' splice donor	(bp)	3' splice acceptor
1	85+5'UTR	ACCTCAGAAGgtgcatcctt	701	cacccctt agGTTGCCAGAT
2	204	AGCCGCTGTGgtgagtagcc	752	ctctccacagTCCGAATCTG
3	186	CACTGCCAGGgtgaggggaa	1351	ctgctgccagTGAATCGAAC
4	21	CCACACTCAGgtgagatgag	1440	ccctcctcagAACGGCGCGC
1b	306	CGACAGCAAGgtagccctgg	3460	tgtctttcagTGTGATCCAG
5	161	CGACAGCAAGgtagccctgg	3460	tgtctttcagTGTGATCCAG
6	135	CCTCATTGTGgtaagcaggg	414	tgccccacagCTTTCCTATG
7	171	CTTCACTTCGgtgaggaggg	1099	cccacccaagACTCTGGACG
8	102	AAACCTGAAGgtcagatggc	286	tctgttgtagCGCCAGGATG
9	66	TTTTTGTGAGgtggagttgg	459	ctgtattcagGTGTACAAGG
10	192	TTCCAACATGgtgagagtgt	7616	tcctcccaagACATCCCAGG
11	243	GGGTGTCTCTgtgagttaaa	625	cttccctcagGGCCATGTGG
12	64	CAGCTTCAGGgttagtacag	1549	gtctgcccagGTGGCAGCTA
13	78	AAATGGATTGgtgagtggat	1544	tgccatgcagGAGGGTCCCC
14	151	CACATGTCCGgtaagtttct	495	tctgccctagTTATATCCAG
15	133	CGTCTGCCAGgtgaggaggt	1382	tcctttctagGCCCGCCTCT
16	117	GTGGAGGAAGgtgagctgct	97	atttatccagACTCTGGAAC
17	108	GACCATTGAGgtaccactgg	412	tctgccctagACATTTGCCA
18	94	ATGGCTTGGTgtgtgggatg	692	gccatcctagGCATTTTCTA
19	128	CAATGTGGCAgtgagcactg	580	ttccctctagGTCCTGTGCC
20	129	TGTGCCCAAGgtaaggatct	222	ttgcccccagATGCGCAGGC
21	144	CATTGCTGAGgtgcgggggt	777	cttcctccagAAAGAGGAGC
22	174+3'UTR			

Table 7. The splice junction that correspond to the consensus sequence

5' splice donor

General consensus	A	G	g	t	a	a	g	t						
Local consensus	T/C	A/G	A	G	G	T	A/G	A	G	N	N	T/G	A/G	T/G
A	1	9	11	1	0	0	6	15	2	4	7	2	5	2
T	9	3	4	2	0	23	1	1	0	6	6	7	3	8
G	4	7	3	19	23	0	14	3	17	7	5	10	10	10
C	8	3	4	0	0	0	0	3	3	5	4	3	4	2

3' splice acceptor

General consensus	y	y	y	y	y	y	n	c	a	g	G			
Local consensus	T/C	T/C	T/C	T/C	T/C	T/C	T/C	T/C	A	G	N	N	N	N
A	0	1	0	2	3	0	1	2	22	0	7	3	6	1
T	11	9	8	7	6	9	6	7	0	0	5	7	6	6
G	2	4	3	3	1	1	2	0	0	22	8	5	4	8
C	8	8	11	10	12	12	12	13	0	0	2	7	6	7

The numbers indicate number of individual nucleotides at the splice junction that correspond to the consensus sequence.

Table 8.

Num. of repeat (CA)n	9	18	19	20	21	22	23	24
Frequencies	2	1	2	6	19	3	2	1
n=36	(5.6%)	(2.8%)	(5.6%)	(16.7%)	(52.7%)	(8.3%)	(5.6%)	(2.8%)

Table 9. Linkage similarity of genes on 6p21.3 - 6p22.1 with those on chromosomes 7, 11, and 17.

Chr. 6		Chr. 7		Chr. 11		Chr.17	
FAT11 3 OLF gene	6p21.3	OLF3 (1 gene)	7p35	OLF1 (1 gene)	11p15.5	OR17 (16 gene)	17p13.3
TRE17 like						TRE17	17p12-p21
MEA11 like	6p21.3	MEA11	7q				
HSRAS like	6p21.3			HSRAS	11p11-12		
RCK like	6p21.3-22.1			RCK/RNA helicase	11q23.3		
RFP	6p21.3-22.1			SSA/Ro	11p15.5-p15.3		

Acknowledgments

I wish to express my sincere gratitude to Prof. Toshimichi Ikemura for guidance and encouragement during all the stages of this work. I thank Dr. Toyoaki Tenzen for collaboration and discussions, and Dr. Ken-ichi Matsumoto, Prof. Naruya Saitou and other members in the Division of Evolutionary Genetics for their encouragements and helpful suggestions. I also express my sincere gratitude to Prof. Hidetoshi Inoko of Tokai Univ. for collaboration and encouragement, and thank Dr. Asako Ando of Tokai Univ., Prof. Katsuzumi Okumura of Mie Univ. for collaboration and discussions. I also thank Mrs. Y. Miyauchi and K. Suzuki for their excellent technical assistance. I acknowledge Mr. Petur. H. Petersen and Dr. Silvana Gaudieri for critical reading of the manuscript. I used computers, programs, and databases of the DDBJ.