Volume 322, number 2, 181–185 FEBS 12424 May 1993

Expression of a novel human homeobox-containing gene that maps to chromosome 7q36.1 in hematopoietic cells

Toshikazu Matsui[a,b], Momoki Hirai[c], Machiko Wakita[b], Masami Hirano[b] and Yoshikazu Kurosawa[a]

[a]*Institute for Comprehensive Medical Science and* [b]*Department of Internal Medicine, Fujita Health University, Toyoake, Aichi 470-11, Japan* and [c]*Department of Anthropology, Faculty of Science, The University of Tokyo, Bunkyo-ku, Tokyo 113, Japan*

Received 27 February 1993

A homeobox is a DNA sequence of 180 base pairs that encodes a DNA-binding domain known as a homeodomain. The polymerase chain reaction (PCR) has been used to prepare probes of homeobox-containing genes. We cloned and sequenced the amplified products of PCR that was performed with human genomic DNA and two primers that correspond to well-conserved regions in homeoboxes. Fifteen kinds of homeobox gene were identified and 13 of them were assigned to *HOX* genes that have already been reported. Two others represented novel homeobox genes and one of them, *GBX1*, was mapped to chromosome 7q36.1 by fluorescence in situ hybridization. Northern hybridization of mRNA for various kinds of hematopoietic cell showed that the newly identified *GBX1* gene is expressed in K562 cells and Daudi cells.

Polymerase chain reaction; Fluorescence in situ hybridization; Hematopoiesis

1. INTRODUCTION

The homoebox was first identified as a common sequence in several homeotic and segmental genes of *Drosophila*, such as *Antp*, *Ubx* and *Ftz* [1]. More than 30 genes involved in the developmental control of *Drosophila* have so far been shown to contain homeobox sequences [2]. The human homologues of most *Drosophila* genes have been isolated by cross-hybridization with *Drosophila* probes and 'chromosome walking' [3]. *Antp*-class genes seem to form the largest family of homeobox genes in the human genome and they are reportedly located in four major complexes: the *HOXA* complex located on chromosome 7 at 7p15-p14 encodes 11 homeobox genes; the *HOXB* complex located at 17q21-q22 encodes 9 homeobox genes; the *HOXC* complex located at 12q13 encodes 9 homeobox genes; and the *HOXD* complex located at 2q31-q37 encodes 9 homeobox genes [4–6]. The 4 *HOX* complexes appear to have been created by multiplication of a primordial gene cluster that corresponds to the *bithorax* complex and the *antennapedia* complex in *Drosophila* [7–9]. Other than those of the *Antp* class, the homeodomains encoded by homeoboxes are found in a large number of transcription factors, and they form a major set of structural motifs that can bind to DNA in a sequence-specific manner [10,11]. In the present study, we attempted to prepare probes of homeobox genes using the polymerase chain reaction (PCR) [12]. Homeodomains display a helix-turn-helix structure and are best conserved in the C-terminal third (helix 3 region) of the domains [2]. Several positions in the helix 1 region are also well conserved [2]. We chose two blocks of amino acid sequences: ELEKEF from the 15th to the 20th position; and WFQNRR from the 48th to the 53rd position as sources of primers. This strategy has already been used for the identification of homeobox genes in various organisms [13–17]. Nucleotide sequence analysis of the amplified products revealed a novel class of homeobox genes in human genomic DNA.

Correspondence address: Y. Kurosawa, Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Aichi 470-11, Japan. Fax: (81) (562) 93-8835.

Abbreviations: FISH, fluorescence in situ hybridization; PCR, polymerase chain reaction.

The sequences reported in this paper have been deposited in the EMBL database (accession nos. X59518, X59519) and in the GenBank database (accession no. L11239).

2. MATERIALS AND METHODS

Human genomic DNA was prepared from human placenta by the published procedure [18]. Two kinds of primer were synthesized chemically with a DNA synthesizer (Applied Biosystems): primer I, GG*GGATCC*GARYTRGARAARGARTT, where Y denotes T or C and R denotes G or A (this base sequence corresponds to the *Bam*HI site, which is underlined, and to the amino acid sequence ELEKEF); and primer II, GG*GTCGAC*YCKYCKRTTYTGRAACCA, where K denotes G or T (this sequence corresponds to the *Sal*I site, which is underlined, and to the amino acid sequence WFQNRR). PCR was performed as described elsewhere [12]. In brief, a 100 μl reaction mixture contained 1 μg of genomic DNA, 2 μM each primer, 200 μM each dNTP, and 2.5 U DNA polymerase from *Thermus aquaticus* (Takara Co.) in 10 mM Tris-HCl (pH 8.4), 50 mM KCl, 2.5 mM $MgCl_2$ and 200 μg/ml gelatin. The mixture was subjected to 33 amplification cycles in a Perkin-Elmer/Cetus Thermocycler: 1 min at 94°C,

2 min at 55°C and 1 min at 74°C. After fractionation by electrophoresis in a 5% agarose gel, the amplified DNAs of around 130 nucleotides in length were digested with both *Bam*HI and *Sal*I and cloned into Bluescript vectors (Stratagene). Nucleotide sequences were determined by the dideoxy chain termination method [19]. Southern hybridization was carried out under non-stringent conditions as described by McGinnis et al. [3]. A genomic library was constructed by partial digestion with *Sau*3A and subsequent ligation with λdash vectors. The library was screened by the Benton–Davis method [20]. After inserts had been separated from plasmid DNA by agarose gel electrophoresis, the probes were labeled by the multi-prime labeling system. The following cells were used for Northern hybridization analysis: K562, Daudi, MOLT4, MOLT3, CCRF-CEM, NALM8, HL60 and Ly16. After cells had been grown in RPMI1640 medium supplemented with 5% FCS, RNAs were extracted by the guanidinium thiocyanate method [20] and poly(A)-containing RNAs were prepared by column chromatography on oligo(dT) cellulose [21]. Northern blots were prepared by the standard method, as described elsewhere [21]. Fluorescence in situ hybridization (FISH) was carried out as described previously [22,23].

3. RESULTS

3.1. *Identification of human homeobox-containing genes by PCR*

Human genomic DNA was subjected to amplification by PCR with the two primers described in section 2. The amplified products were separated by electrophoresis on an agarose gel. One major band of around 130 nucleotides was observed. The fragments were eluted from the gel and cloned in the vector. Nucleotide sequences of 28 independent clones were determined. All of them contained fragments of identical size sandwiched between the two primers. Fifteen different sequences were identified and compared with those reported in the literature. Thirteen of them were assigned to *HOX* genes that have already been reported [4–6] (the numbers in parentheses indicate the number of isolated clones with each sequence): *HOXA1* (2); *HOXA5* (4); *HOXA6* (2); *HOXA7* (1); *HOXB1* (3); *HOXB5* (1); *HOXB6* (1); *HOXB7* (1); *HOXB8* (2); *HOXB9* (1); *HOXC4* (2); *HOXC8* (1); and *HOXD1* (1). Although a few discrepancies were observed between the nucleotide sequences determined in the present study and those reported previously, all of them were located at the third letter in each codon and were probably due to polymorphism. Nucleotide sequences of two other genes, *GBX1* and *GBX2*, are indicated in Fig. 1. Although the nucleotide sequences of *GBX1* and *GBX2* were only 84% identical, the amino acid sequences deduced from them were 100% identical to each other. These sequences did not correspond to any of the homeobox genes that have been identified to date in the human genome, however, when the amino acid sequence was compared with those encoded by all the known homeobox genes in various organisms, it was found to correspond to that encoded by *CHox-7*, identified in the chicken [24]. Moreover, amino acid sequences encoded by *MMoxA* and *MMoxB*, which were identified from among the mRNAs of the mouse telencephalon by the same PCR strategy as exploited in this study, were also the same as those encoded by *GBX1* and *GBX2* [15]. In the case of *MMoxA* and *MMoxB*, partial sequences from cysteine at the second position to lysine at the 2nd position from the carboxy-terminal are known [15]. Since the mouse *MMoxB* and *MMoxA* genes have recently been renamed *Gbx1* and *Gbx2*, respectively (Dr. M. Murtha. personal communication), we refer to our human genes as *GBX1* and *GBX2*. A genomic library was constructed from human placental DNA and screened with *GBX1* DNA as probe. One clone, λGBX1-1, which was positive for hybridization with the *GBX1* probe, was isolated and analyzed. Fig. 2 shows the nucleotide sequence of the region that includes the *GBX1* probe-positive sequence. It clearly contains a homeobox sequence. The possible coding region was identified by comparison of the predicted amino acid sequence with that encoded by *CHox-7* [24]. The amino acid sequences of the homeodomains encoded by *GBX1* and *CHox-7* were 100% identical to each other. Downstream of the homeodomain, the first 18 amino acid residues of the two putative proteins are very similar but then the homology suddenly disappears, however, if one nucleotide is deleted from the published sequence of *CHox-7*, the extent of homology again becomes very high. We repeated our sequencing of *GBX1* in both directions and confirmed our sequence. Thus, the differences observed may be due to an error in sequencing of *CHox-7* [24]. Since there is no consensus sequence for splicing before a stop codon appears at the 44th residue downstream from the homeodomain, this stop codon should be the end of the coding region of the *GBX1* gene. The amino acid sequences encoded by *GBX1* and *CHox-7* upstream of the homeodomain are rather different, however, the first eight amino acid residues encoded by *CHox-7*, namely, GPK*G*K*G*KG, are very similar to those encoded by *GBX1*, namely, GPK*P*K*L*KG. This similarity indicates that this region should be included in the coding exon. Since a stop codon is present, in frame, 39 bp upstream of the sequence that encodes GPGA, the 5′ end of this exon was tentatively identified and is indicated in Fig. 2.

```
          H  C  K  K  Y  L  S  L  T  E  R  S  Q  I  A  H  A  L  K  L  S  E  V  Q  V  K  I
GBX1  TCATTGCAAGAAATACCTGAGCTTGACAGAGCGCTCTCAGATCGCCCACGCCCTCAAGCTCAGTGAGGTGCAGGTCAAGATC
GBX2  C--C-----A--G-----CTC------C--------G-------------------A-----C----------G--A---
```

Fig. 1. Nucleotide sequences of the products of PCR. Bars indicate the nucleotides that are the same as those in *GBX1*. The same amino acid sequences are encoded by *GBX1* and *GBX2*.

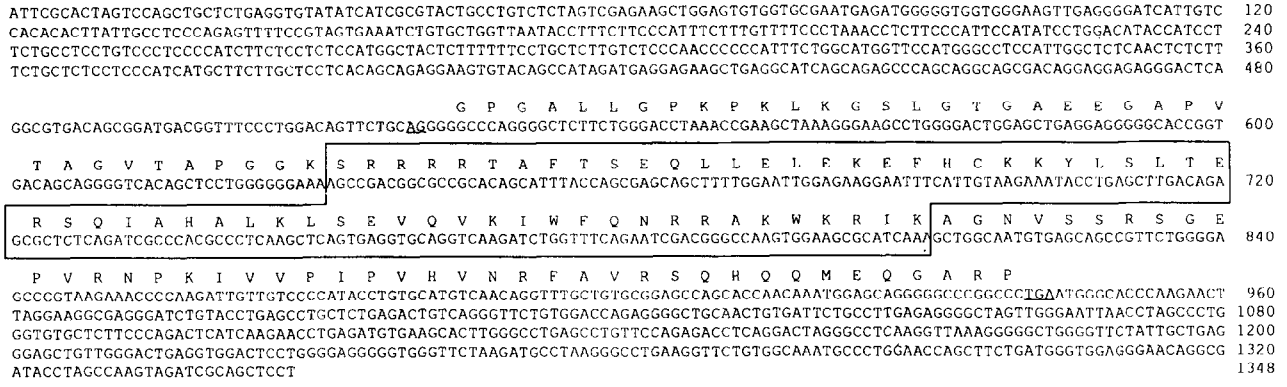

Fig. 2. Nucleotide sequence of λGBX1-1. The homeobox is boxed. The amino acid sequence of a putative exon is also indicated.

3.2. *Localization of* GBX1 *on human chromosome 7 at 7q36.1*

The location of the *GBX1* gene on a human chromosome was determined by the FISH method [22,23]. As shown in Fig. 3, the *GBX1* gene was mapped to chromosome 7q36.1. This result seems to be in good agreement with that found in the mouse, in which the *Gbx1* gene has been mapped to chromosome 5 (Dr. M. Murtha, personal communication) and the mouse homologues of many genes on human chromosome 7 are encoded on mouse chromosome 5 [25–27].

3.3. *Expression of GBX1 in hematopoietic cells*

mRNAs were prepared from various kinds of hematopoietic cell and subjected to Northern hybridization analysis with *GBX1* and *GBX2* as probes. As shown in Fig. 4A, *GBX1* was expressed in both K562 cells and Daudi cells. The size of the transcript was around 2.2 kb. *GBX1* and *GBX2* seem to form a new class of homeodomain that has at least three members, as shown in Fig. 4B. Although the extent of sequence homology between the *GBX1* and *GBX2* probes was 84% and both probes detected common bands of different intensities during Southern hybridization, the *GBX2* probe did not hybridize with any bands during Northern hybridization (data not shown). Therefore, the expressed transcript of 2.2 kb in K562 and Daudi cells should be an authentic transcript of the *GBX1* gene.

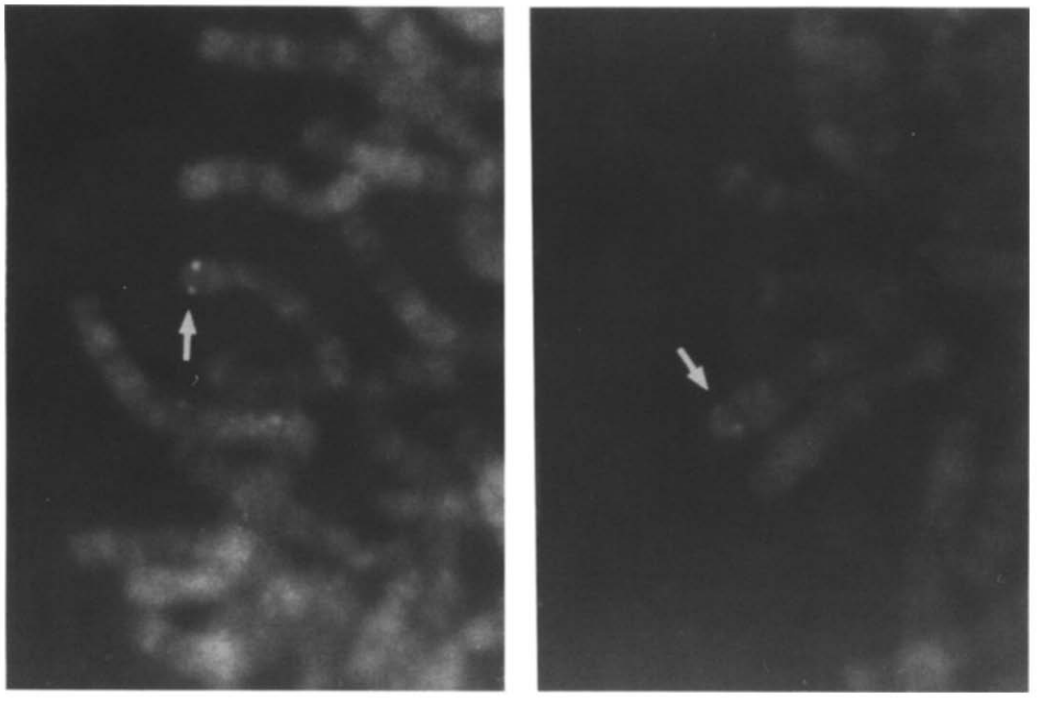

Fig. 3. Results of fluorescence in situ hybridization with the biotinylated human *GBX1* probe and human chromosomes. R-banded metaphases after in situ hybridization are shown. Arrows indicate signals at 7q36.1.

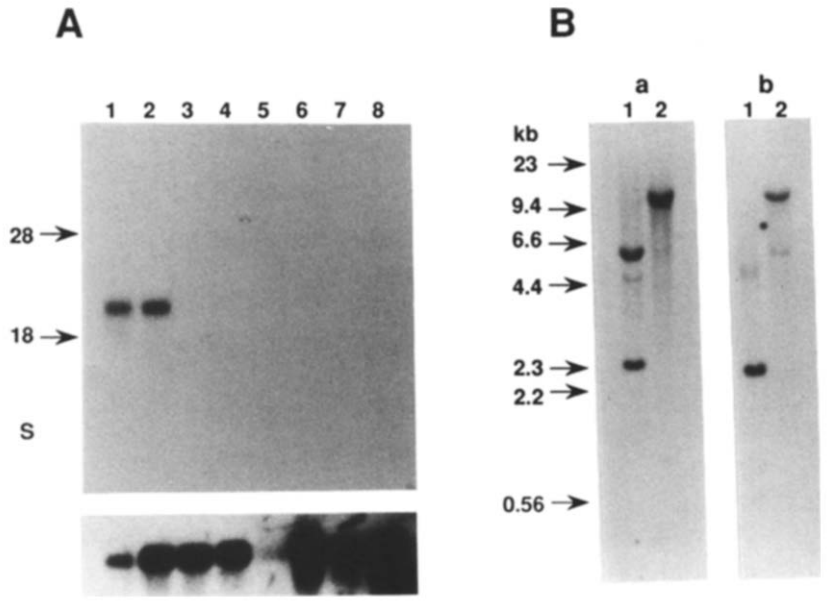

Fig. 4. Expression of the *GBX1* gene in hematopoietic cells. (A) Northern hybridization of various lines of cells with the *GBX1* probe. Lanes: 1, K562; 2, Daudi; 3, Molt4; 4, Molt3; 5, CCRF-CEM; 6, NALM8; 7, HL60; 8, Ly16. Lower panel: a DNA fragment encoding the gene for β-actin was used as a probe [40]. (B) Southern hybridization of human genomic DNA with *GBX1* and *GBX2 probes*. Probes: a, *GBX2*; b, *GBX1*. Restriction enzymes: 1, *Hin*dIII; 2, *Eco*RI.

4. DISCUSSION

In this study, we used the PCR method to prepare probes for homeobox genes in the human genome. Many mammalian homeobox genes have been isolated by cross-hybridization with *Drosophila* probes [3]. Degenerate oligonucleotides corresponding to the highly conserved sequence in the helix 3 region have also been utilized as probes for identification of homeobox genes [28]. Analysis of genes in the vicinity of break points associated with chromosome translocation in leukemic cells has resulted in the identification of novel homeobox genes [29]. The advantage of the PCR method, as adopted in the present study, is that it allows the direct acquisition of sequence information in a short time. We were able to judge whether identified genes were novel or already known. In fact, the *GBX1* and *GBX2* genes identified in the present study are novel homeobox genes in the human genome. Regions of 27 amino acids encoded by the amplified fragments of *GBX1* and *GBX2* are identical to each other. Murtha et al. [15] have already identified the mouse homologues of these two genes. This class of the homeodomain may have one more member, as shown in Fig. 4B. The amino acid sequence of the homeodomain encoded by *GBX1* is 100% identical to that encoded by *CHox-7* and identified in chicken. A high degree of conservation of sequence is also found in the surrounding regions. Thus, this class of homeobox gene appears to have been effectively conserved during evolution.

The *GBX1* gene is located at chromosome 7q36.1. Another homeobox gene, *EN-2*, has been mapped to this region [30,31]. The mouse *En-2* gene was also mapped to mouse chromosome 5 [30,31]. Although the sequence of *GBX2* is very similar to that of *GBX1*, this similarity does not necessarily indicate that the location of *GBX2* is close to that of *GBX1*. The *EN-1* gene, which is homologous to *EN-2*, was mapped to chromosome 2 in human [30] and to chromosome 1 in the mouse [32,33]. Moreover, *PBX-1*, *PBX-2* and *PBX-3*, which have structures very similar to one another, are located at different chromosomal loci [34]. The same is true for the *msh*-like genes, *HOX7.1* and *HOX8.1* [35].

GBX1 was expressed in two kinds of hematopoietic cell, K562 cells and Daudi cells. It has been well established that the homeobox genes play pivotal roles in determining the anterior–posterior body axis of the embryo and in specifying different cell types during animal development [36]. The hematopoietic system is organized in a developmental hierarchy in which stem cells in bone marrow can mature into any of the cells of the hematopoietic lineage. The homeobox genes seem to be strong candidates for genes that determine the lineage commitment within the hematopoietic system. Kongsuwan et al. [37] and Shen et al. [38] analyzed the expression of several *HOX* genes in various cell lines that represented erythroid, myeloid and lymphocyte lineages. Magli et al. [39] systematically examined the expression of all the genes encoded at four *HOX* loci in various cell lines, which included erythroleukemic, promyelocytic and monocytic cells. There are some discrepancies among the reported results. For example, *HoxA7* is expressed in all hematopoietic cells according to Konsuwan et al. [37], but not according to Magli et al. [39].

We repeated similar experiments using *HOXA1*, *HOXA7*, *HOXB1*, *HOXB4*, *HOXB7*, *HOXB8* and *HOXB9* probes. None of the cells tested (the lines are listed in the legend to Fig. 4A) expressed these transcripts, as indicated by Northern hybridization analysis (data not shown). Further experiments are required to establish the possible involvement of homeobox genes in hematopoiesis.

Acknowledgements: We thank Dr. M. Murtha for providing us with information about mouse *Gbx1* and *Gbx2* genes. This work was supported in part by grants from Japan Foundation for Aging and Health; Ministry of Education, Science and Culture in Japan; Fujita Health University.

REFERENCES

[1] McGinnis, W., Levine, M.S., Hafen, E., Kuroiwa, A. and Gehring, W.J. (1984) Nature 308, 428–433.
[2] Scott, M.P., Tamkun, J.W. and Hartzell III, G.W. (1989) Biochim. Biophys. Acta 989, 25–48.
[3] McGinnis, W., Hart, C.P., Gehring, W.J. and Ruddle, F.H. (1984) Cell 38, 675–680.
[4] Acampora, D., D'Esposito, M., Faiella, A., Pannese, M., Migliaccio, E., Morelli, F., Stornaiuolo, A., Nigro, V., Simeone, A. and Boncinelli, E. (1989) Nucleic Acids Res. 17, 10385–10402.
[5] Simeone, A., Acampora, D., Nigro, V., Faiella, A., D'Esposito, M., Stornaiuolo, A., Mavilio, F. and Boncinelli, E. (1991) Mech. Dev. 33, 215–228.
[6] Scott, M.P. (1992) Cell 71, 551–553.
[7] Duboule, D. and Dolle, P. (1989) EMBO J. 8, 1497–1505.
[8] Graham, A., Papalopulu, N. and Krumlauf, R. (1989) Cell 57, 367–378.
[9] Kappen, C., Schughart, K. and Ruddle, F.H. (1989) Proc. Natl. Acad. Sci. USA 86, 5459 5463.
[10] Desplan, C., Theis, J. and O'Farrell, P.H. (1988) Cell 54, 1081–1090.
[11] Pabo, C.O. and Sauer, R.T. (1992) Annu. Rev. Biochem. 61, 1053–1093.
[12] Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Science 239, 487–491.
[13] Kamb, A., Weir, M., Rudy, B., Varmus, H. and Kenyon, C. (1989) Proc. Natl. Acad. Sci. USA 86, 4372–4376.
[14] Frohman, M.A., Boyle, M. and Martin, G.R. (1990) Development 110, 589–607.
[15] Murtha, M.T., Leckman, J.F. and Ruddle, F.H. (1991) Proc. Natl. Acad. Sci. USA 88, 10711–10715.
[16] Naito, M., Kohara, Y. and Kurosawa, Y. (1992) Nucleic Acids Res. 20, 2967–2969.
[17] Shenk, M.A., Bode, H.R. and Steele, R.E. (1993) Development (in press).
[18] Gross-Bellard, M., Oudet, P. and Chambon, P. (1973) Eur. J. Biochem. 36, 32–38.
[19] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[20] Benton, W.D. and Davis, R.W. (1977) Science 196, 180–182.
[21] Sambrook, J., Fritch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Lab., Cold Spring Harbor, NY.
[22] Takahashi, E., Hori, T., O'Connell, P., Leppert, M. and White, R. (1990) Human Genet. 86. 14–16.
[23] Ichihara, Y., Hirai, M. and Kurosawa, Y. (1992) Immunol. Lett. 33, 277–284.
[24] Fainsod, A. and Gruenbaum, Y. (1989) FEBS Lett. 250, 381–385.
[25] O'Brien, S.J. and Graves, J.A.M. (1991) Cytogen. Cell Genet. 58, 1124–1151.
[26] Nadeau, J.H., Davisson, M.T., Doolittle, D.P., Grant, P., Hillyard, A.L., Kosowsky, M.R. and Roderick, T.H. (1992) Mammalian Genome 3, 480–536.
[27] O'Brien, S.J., Womack, J.E., Lyons, L.A., Moore, K.J., Jenkins, N.A. and Copeland, N.G. (1993) Nature Genet. 3, 103–112.
[28] Bürglin, T.R., Finney, M., Coulson, A. and Ruvkun, G. (1989) Nature 341, 239–243.
[29] Kamps, M.P., Murre, C., Sun, X.H. and Baltimore, D. (1990) Cell 60, 547–555.
[30] Logan, C., Willard, H.F., Ronamens, J.M. and Joyner, A.L. (1989) Genomics 4, 206–209.
[31] Poole, S.J., Law, M.L., Kao, F.T. and Lau, Y.F. (1989) Genomics 4, 225–231.
[32] Joyner, A.L. and Martin, G.R. (1987) Genes Dev. 1, 29–38.
[33] Martin, G.R., Richman, M., Reinsch, S., Nadeau, J.H. and Joyner, A. (1990) Genomics 6, 302–308.
[34] Monica, K., Galili, N., Nourse, J., Saltman, D. and Cleary, M.L. (1991) Mol. Cell Biol. 11, 6149–6157.
[35] Monaghan, A.P., Davidson, D.R., Sime, C., Graham, E., Baldock, R., Bhattacharya, S.S. and Hill, R.E. (1991) Development 112, 1053–1061.
[36] McGinnis, W. and Krumlauf, R. (1992) Cell 68, 283–302.
[37] Kongsuwan, K., Webb, E., Housiaux, P. and Adams, J.M. (1988) EMBO J. 7, 2131–2138.
[38] Shen, W.-F., Largman, C., Lowney, P., Corral, J.C., Detmer, K., Hauser, C.A., Simonitch, T.A., Hack, F.M. and Lawrence, H.J. (1989) Proc. Natl. Acad. Sci. USA 86, 8536–8540.
[39] Magli, M.C., Barba, P., Caletti, A., De Vita, G., Cillo, C. and Boncinelli, E. (1991) Proc. Natl. Acad. Sci. USA 88, 6348–6352.
[40] Ponte, P., Ng, S.-Y., Engel, J., Gunning, P. and Kedes, L. (1984) Nucl. Acids Res. 12, 1687–1696.