Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 960–963, February 1990
Genetics

A cluster of hematopoietic serine protease genes is found on the same chromosomal band as the human α/δ T-cell receptor locus

(cathepsin G/cytotoxic T cell/*TCRA/TCRD* locus/*CTSG* gene)

ROBIN D. HANSON*, PATRICIA A. HOHN*, NICHOLAS C. POPESCU†, AND TIMOTHY J. LEY*‡

*Division of Hematology/Oncology, Departments of Medicine and Genetics, Jewish Hospital at Washington University Medical Center, 216 South Kingshighway Boulevard, St. Louis, MO 63110; and †Laboratory of Biology, National Cancer Institute, Building 37, Room 2A19, Bethesda, MD 20892

Communicated by Philip W. Majerus, November 20, 1989 (received for review October 10, 1989)

ABSTRACT **The chymotrypsin-like family of serine protease genes includes several members that are expressed exclusively in subsets of hematopoietic cells. For example, human neutrophil elastase and cathepsin G are expressed only in myelomonocytic precursors, and cytotoxic-T-cell serine proteases are found only in cytotoxic lymphocytes. We have used a cathepsin G cDNA probe to clone two cathepsin G-like genes (designated CGL-1 and CGL-2) from a human genomic library. We have determined that CGL-1 is identical to a previously identified gene (known as CCPI, CTLA I, or cytotoxic serine protease B) that is expressed only in activated cytotoxic T lymphocytes. We show here that cathepsin G, CGL-1, and CGL-2 are linked on an ≈50-kilobase locus found on human chromosome 14 at band q11.2. This gene cluster maps to the same chromosomal band as the α and δ T-cell receptor genes; this region is involved in most chromosomal translocations and inversions that are specifically associated with T-cell malignancies.**

During the past several years, a number of investigations have begun to define a group of highly related hematopoietic serine protease genes. Although the organization of these genes suggests that they arose from a single primordial gene (1), the expression of the various members of this gene family is usually restricted to a single lineage of hematopoietic cells at specific times during development. For example, human cathepsin G and neutrophil elastase are expressed in myelomonocytic precursors (2, 3), rat mast-cell proteases I and II are expressed in mast cells (4, 5), and the murine cytotoxic-T-cell serine proteases CCPI, CCPII, granzymes D, E, and F, and Hanukah factor are found only in cytotoxic lymphocytes (6–8).

We performed the present study to define a family of genes related to human cathepsin G, a neutral serine protease found in the azurophil granules of myelomonocytic cells (9–11). We recently described the genomic structure of this gene and localized it to human chromosome 14 at band q11.2 (12). In this study, we determined that a cytotoxic-T-cell serine protease gene and another closely related gene (or pseudogene) are tightly linked to the cathepsin G gene, forming an hematopoietic serine protease gene cluster.

MATERIALS AND METHODS

Library Screening. A human genomic library in λ-Fix (Stratagene) was screened with a random-primer-labeled probe derived from cathepsin G cDNA (12). Hybridization was performed at 42°C in 50% formamide/10% dextran sulfate/1 M NaCl/1% SDS containing denatured sonicated salmon sperm DNA at 100 μg/ml. Two washes were performed at 42°C in 2× SSC/0.1% SDS, followed by two washes with 0.1× SSC/0.1% SDS at room temperature. (SSC is 0.15 M NaCl/0.015 M sodium citrate, pH 7.)

A human cosmid library in the vector pCK (13) was screened with the unique cathepsin G, CGL-1, and CGL-2 probes illustrated in Fig. 4. The cathepsin G probe is a 0.9-kilobase (kb) *Eco*RI–*Hin*dIII fragment, the CGL-1 probe is a 1.8-kb *Eco*RI fragment, and the CGL-2 probe is a 1.2-kb *Bam*HI–*Pst* I fragment. Hybridization was performed at 42°C in 50% formamide/12.5% dextran sulfate/5× SSC/10 mM Tris·HCl, pH 8.0/1× Denhardt's solution/0.1% SDS. (Denhardt's solution is 0.02% polyvinylpyrrolidone/0.02% Ficoll/0.02% bovine serum albumin.) Final washes were performed at 65°C in 0.1× SSC/0.1% SDS.

Southern Blot Analysis. Conventional Southern blot analysis was performed as described (12). DNA was prepared from human peripheral blood mononuclear cells, digested with restriction enzyme, electrophoresed in agarose gel, and transferred to GeneScreen*Plus* (New England Nuclear). Hybridization was performed at 42°C in 50% formamide/10% dextran sulfate/1 M NaCl/1% SDS containing denatured sonicated salmon sperm DNA at 100 μg/ml. Final washes were performed at 65°C in 0.1× SSC/0.1% SDS.

For pulsed-field Southern blot analysis, K562 cell plugs were prepared, digested, and electrophoresed at 160 V (6 V/cm) with an 80-sec constant switch interval for 24 hr (14). DNA was transferred to nylon-backed nitrocellulose (Nitro Plus 2000, Micron Separations, Westboro, MA) (15). Samples were run in duplicate and the membrane was cut in half after transfer. Both membranes were hybridized consecutively with the cathepsin G, CGL-1, and CGL-2 probes described above. Hybridization was performed at 42°C in 50% formamide/10% dextran sulfate/4× SSC/0.8× Denhardt's solution/7 mM Tris·HCl, pH 7.6, containing denatured sonicated salmon sperm DNA at 100 μg/ml. Final washes were performed at 65°C in 0.1× SSC/0.1% SDS. Membranes were stripped between each probe hybridization and autoradiographed to ensure complete removal of the previous probe.

Bacteriophage and cosmid clones were digested with appropriate restriction enzymes, electrophoresed, and blotted as described above. Unique probes from various regions of the locus were prepared from low-melting-point agarose gels, labeled by the random-primer method, and hybridized as described above.

Chromosomal *in Situ* Hybridization. Human metaphase and prometaphase preparations from methotrexate-synchronized peripheral blood lymphocyte cultures were treated with RNase A for 1 hr at 37°C and denatured in 70% formamide at 70°C for 2 min. The CGL-1 probe was nick-translated with all four [^{3}H]dNTPs (Amersham) to a specific activity of 3.8×10^7 cpm/μg of DNA. Hybridizations, washes, and counterstaining were performed as described (16).

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

‡To whom reprint requests should be addressed.

RESULTS

Identification of Cathepsin G-Like (CGL) Genes. We previously showed that a human cathepsin G cDNA probe cross-hybridizes at moderate stringency with at least two other genomic DNA fragments (12). We therefore used a cathepsin G cDNA fragment to clone these related sequences from a human genomic library in bacteriophage λ-Fix. Ten clones were analyzed and found to correspond to three nonoverlapping loci, including the cathepsin G gene itself. Each locus contained a single region with cathepsin G homology. The two cathepsin G-like loci were designated CGL-1 and CGL-2. DNA probes derived from the 5′ ends of the cathepsin G, CGL-1, and CGL-2 genes (shown as black bars in Fig. 4) were used to probe Southern blots containing human genomic DNA cut to completion with *Bam*HI, *Bgl* II, or *Eco*RI (Fig. 1). No cross-hybridization among the three loci was observed with the unique 5′ probes, and no fragment hybridized with more than one probe. The pattern of hybridization for the CGL-1 probe was identical to the pattern seen with cDNA probes encoding a cytotoxic-T-cell serine protease known as CSP-B (17, 18), the product of the human homologue of the murine CCPI gene (also known as granzyme B or CTLA-I) (6, 19, 20).

mRNA derived from the human CSP-B (CGL-1) gene has been detected only in peripheral blood lymphocytes stimulated with phytohemagglutinin interleukin 2, or *Staphylococcus aureus* enterotoxin A (17, 18). To further confirm the identity of CGL-1, we hybridized the 5′ CGL-1 probe (see Fig. 4) with RNA derived from untreated and phytohemagglutinin-stimulated human peripheral blood lymphocytes, the myelomonocytic cell lines U-937 and PLB-985 (21), HeLa cells, and K562 erythroleukemia cells in a standard Northern blot analysis (data not shown). The CGL-1 probe detected a 0.9-kb mRNA only in the sample derived from phytohemagglutinin-stimulated peripheral blood mononuclear cells (18) even though all lanes contained similar amounts of hybridizable β-actin mRNA (22) (data not shown). Finally, sequencing of CGL-1 from position −200 to position +111 (with respect to the transcription initiation site) revealed identity with the sequence of CSP-B (23) (data not shown).

The identity of CGL-2 has not been established. The genomic fragments detected by the 5′ CGL-2 probe do not resemble published patterns for previously cloned serine proteases. Hybridization and initial sequencing studies suggest that the gene is highly related to both cathepsin G and CGL-1; further experiments will be required to determine whether CGL-2 is a pseudogene or yet another functional serine protease with a restricted pattern of expression.

Chromosomal Localization of CGL-1. We previously demonstrated that the human cathepsin G gene is found on chromosome 14 at band q11.2 (12). To determine whether CGL-1 is located at the same site, we radiolabeled the specific CGL-1 probe (shown in Fig. 4) and performed chromosomal *in situ* hybridization. One hundred two pro- and mid-metaphase chromosomes were informatively labeled; 48 of 176 evaluable grains were localized on chromosome 14. Twenty-nine grains clustered at band q11.2 (Fig. 2). These data are consistent with a previous study (24) that localized CTLA-1 (CGL-1) to human chromosome 14 at bands q11–q12. This is also the chromosomal location of the genes encoding the α and δ chains of the T-cell antigen receptor (25, 26).

Linkage of the Cathepsin G, CGL-1, and CGL-2 Genes. We next assessed the linkage among cathepsin G, CGL-1, and CGL-2 by hybridizing the same unique, non-cross-hybridizing probes shown in Fig. 4 with large genomic DNA fragments resolved by pulsed-field gel electrophoresis. All three probes hybridized with identically sized *Not* I, *Sal* I, and *Bss*HII DNA fragments derived from K562 cells (Fig. 3). Hybridization of each of these unique probes with a 200-kb *Bss*HII fragment suggested that the three genes are closely linked.

To complete the physical linkage studies of the cathepsin G, CGL-1, and CGL-2 genes, we screened a human cosmid library with probes for all three genes. Four hybridizing cosmids were identified and characterized. As shown in Fig.

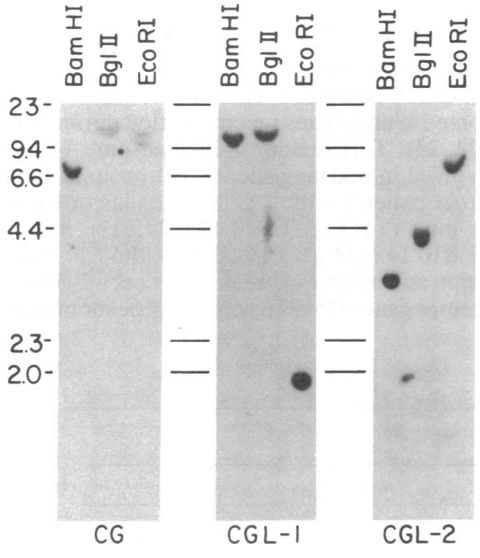

FIG. 1. Southern blot analysis of genomic human DNA with cathepsin G, CGL-1, and CGL-2 probes derived from the 5′ regions of each gene (black boxes beneath each gene in Fig. 4). The cathepsin G (CG) probe is a 0.9-kb *Hin*dIII–*Eco*RI fragment, the CGL-1 probe is a 1.8-kb *Eco*RI fragment, and the CGL-2 probe is a 1.2-kb *Bam*HI–*Pst* I fragment. Size markers (kb) are shown at left. Note that each probe is unique in the human genome and that each hybridizes with a different set of DNA fragments. Two bands are seen with *Eco*RI-cleaved DNA hybridized with the cathepsin G probe; this represents an allelic polymorphism that was not detected in other human DNA samples (data not shown).

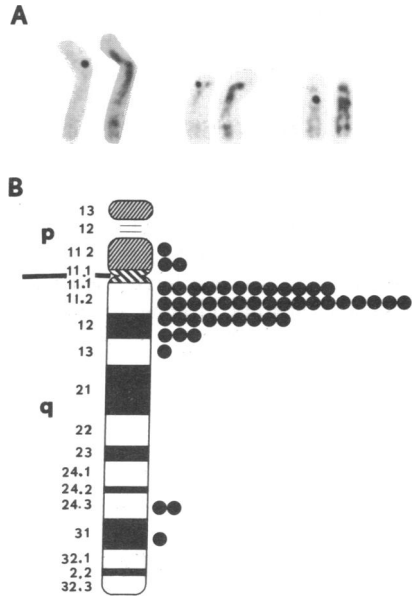

FIG. 2. Chromosomal localization of the CGL-1 gene. (*A*) Three representative pairs of photographs of chromosome 14 with different degrees of retraction. The left member of each pair is a photograph of the *in situ* hybridization; a single silver grain is seen on the long arm near the centromere. The right member of each pair is the same chromosome photographed after G-banding. Left pair, prometaphase; middle pair, early metaphase; right pair, mid-metaphase. (*B*) Distribution of autoradiograph silver grains on chromosome 14; the largest accumulation of grains is on band q11.2.

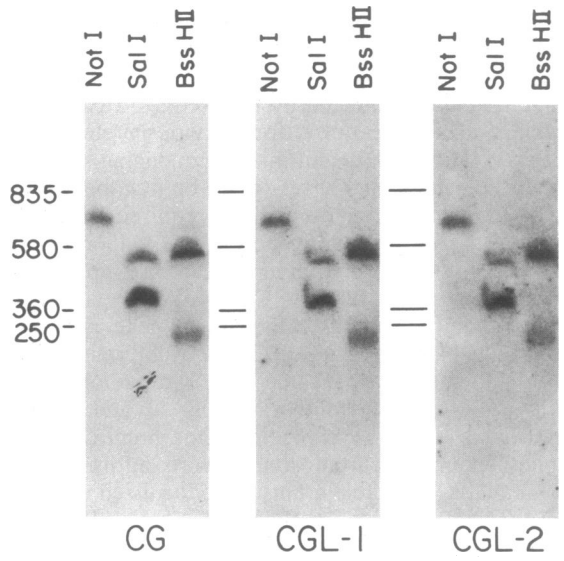

FIG. 3. Pulsed-field Southern blot analysis with the cathepsin G (CG), CGL-1, and CGL-2 probes. High molecular weight K562 cell DNA was digested with the indicated enzymes, separated by pulsed-field gel electrophoresis (clamped homogeneous electric field), and blotted onto nitrocellulose. Probes are described in Fig. 1. Size markers (kb) are shown at left. Note that all three probes hybridize with identical *Not* I, *Sal* I, and *Bss*HII fragments, suggesting that these three fragments are ≤200 kb apart in the human genome. The presence of two hybridizing bands with *Sal* I and *Bss*HII digests probably represents partially digested DNA, perhaps due to partial methylation of DNA in one of the recognition sites.

4, two cosmids, C2.10 and C3.23, contain all or part of both CGL-1 and CGL-2. C6.32 contains CGL-2 and the first exon of the cathepsin G gene, and C5.18 overlaps the 3′ end of C6.32 and extends downstream from the cathepsin G gene. CGL-1 and CGL-2 are separated by about 21 kb, and the cathepsin G gene is about 31 kb downstream of CGL-2. All three genes are in the same 5′ → 3′ orientation (the orientation of CGL-2 was predicted by hybridization with exon-specific probes generated from the cathepsin G gene).

DISCUSSION

We have cloned and identified two genes that are highly related to cathepsin G. We localized one of these genes, CGL-1, to chromosome 14, band q11.2, the same chromosomal location as the cathepsin G gene and the α/δ T-cell receptor locus (12, 25, 26). We have used pulsed-field gel electrophoresis and cosmid cloning to establish physical linkage of the cathepsin G, CGL-1, and CGL-2 genes. These three genes are contained in an hematopoietic serine protease gene cluster encompassing ≈50 kb of DNA. We have determined with restriction mapping, sequencing, and expression analysis that CGL-1 is identical to a previously described serine protease gene that is expressed only in cytotoxic T lymphocytes.

The murine homologue of CGL-1 has been mapped to mouse chromosome 14, band D (20), also the location of the murine α and δ T-cell receptor genes. The conservation of the linkage between the α/δ T-cell receptor locus and the hematopoietic serine protease gene cluster of the mouse and human genomes may indicate regulatory interactions between these two loci. Since the α/δ T-cell receptor locus is large and complex, we have not yet been able to link this locus and the hematopoietic serine protease gene cluster. A recent study of the inv(14)(q11;q32) chromosomal inversion present in SUP-T1 cells suggested that the CTLA-I/CGL-1 gene is telomeric to the α/δ T-cell receptor locus (24); linkage of the CTLA-I and α/δ T-cell receptor loci could not be established by pulsed-field gel analysis of 50- to 700-kb DNA fragments (24).

Since CGL-1/CSP-B is expressed only in activated cytotoxic lymphocytes (6, 17–20), cathepsin G is expressed only during the promyelocyte/promonocyte stage of development (2, 9, 27), and the α/δ T-cell receptor genes first become active during early T-cell ontogeny (25, 26, 28, 29), this region of the genome could be coordinately organized during very early stages of hematopoietic development. A potentially similar system has been described for the human β-globin gene cluster (30–34). A locus activating region (LAR) 5′ to the cluster of embryonic, fetal, and adult β-globin-like genes (a region that spans ≈100 kb of DNA) appears to mediate organization of the locus into a transcriptionally accessible domain in erythroid cells at all stages of development (30–34). The LAR functions in combination with globin gene promoters and enhancers to direct high-level, tissue-specific expression of the different globin genes at different times of erythroid development. It is possible that a related leukocyte LAR element may exist near the serine protease gene cluster or the α/δ T-cell receptor locus.

The α/δ T-cell receptor locus, also found at band q11 of chromosome 14, undergoes recombination during T-cell maturation (24, 25). This region of chromosome 14 is nonrandomly involved in rearrangements that occur in tumor cells derived from patients with T-cell leukemias or lymphomas, including inv(14)(q11;q32), t(8:14)(q24;q11), t(11;14)(p13;q11), and t(10;14)(q24;q11) (34–36). Although many 14q11 translocation and inversion breakpoints occur within the α/δ T-cell receptor genes (35–37), some may occur outside of this

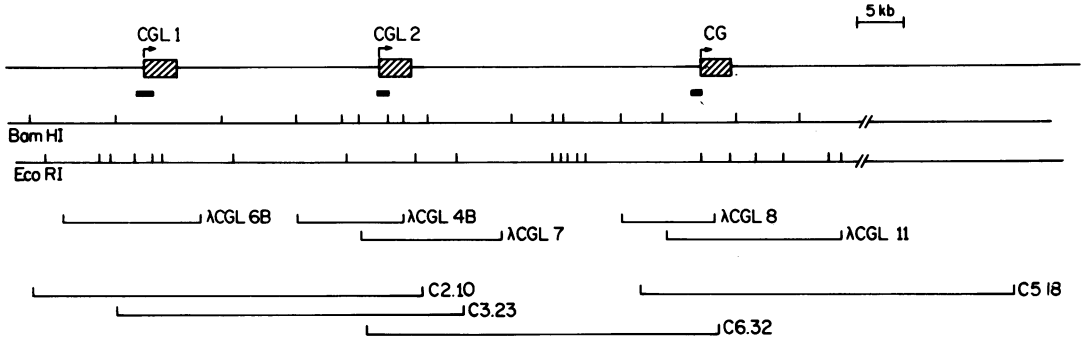

FIG. 4. Map of the serine protease gene cluster. The cathepsin G (CG), CGL-1, and CGL-2 genes are shown as hatched boxes with arrows indicating the transcriptional orientation of each gene. The locations of probes used in this study are indicated by the black bars below each gene. The positions of *Bam*HI and *Eco*RI restriction sites are shown. Relative positions of the λ-Fix clones (λCGL 4B, 6B, 7, 8, and 11) and cosmid clones (C2.10, C3.23, C5.18, and C6.32) are shown. The predicted orientation of CGL-2 was determined by cross-hybridization with genomic fragments containing exon 1, exons 2–4, or exon 5 of the cathepsin G gene.

locus. For example, a protooncogene translocated into the serine protease gene cluster could potentially be disregulated by a serine protease gene regulatory element. Conversely, disruption of the T-cell receptor locus could disregulate serine protease gene expression and possibly alter normal cell growth. One recent study has demonstrated that a cDNA fragment derived from the murine cytotoxic-T-cell serine protease granzyme A encodes a secreted protein that has mitogenic activity (38). Furthermore, we have detected low levels of cathepsin G mRNA in the 8402 T-cell leukemia cell line, which contains a t(11;14)(p13;q11) translocation that has been shown to break within the δ T-cell receptor gene (36) (data not shown). Detailed analysis of the structure and function of the cathepsin G, CGL-1, and CGL-2 genes in additional cell lines containing chromosomes with breakpoints at 14q11 may provide further clues regarding the regulation of this locus and its possible involvement in T-cell malignancies.

We thank Drs. Gary Silverman and Stan Korsmeyer for assistance with pulsed-field gel analyses and Dr. David Chaplin for assistance with cosmid cloning. Diana Horn provided expert editorial assistance. This work was supported by National Institutes of Health Grants DK38682 and CA49712 to T.J.L.

1. Brenner, S. (1988) *Nature (London)* **334,** 528–530.
2. Hanson, R. D., Connolly, N. L., Burnett, D. B., Campbell, E. J., Senior, R. D. & Ley, T. J. (1990) *J. Biol. Chem.*, in press.
3. Fouret, P., DuBois, R. M., Bernaudin, J. F., Takahashi, H., Ferrans, V. J. & Crystal, R. G. (1989) *J. Exp. Med.* **169,** 833–845.
4. Woodbury, R. G., Grazenski, G. M. & Lagunoff, D. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 2785–2789.
5. Benfey, P. N., Yin, F. H. & Leder, P. (1987) *J. Biol. Chem.* **262,** 5377–5384.
6. Lobe, C. G., Havele, C. & Bleackley, R. C. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 1448–1452.
7. Jenne, D., Rey, C., Haefliger, J., Qiao, B., Broscurth, P. & Tschopp, J. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 4814–4818.
8. Gershenfeld, H. K. & Weissman, I. L. (1986) *Science* **232,** 854–858.
9. Senior, R. M., Campbell, E. J., Landis, J. A., Cox, F. R., Kuhn, C. & Koren, H. S. (1982) *J. Clin. Invest.* **69,** 384–393.
10. Senior, R. M. & Campbell, E. J. (1984) *J. Immunol.* **132,** 2547–2551.
11. Campbell, E. J., Silverman, E. K. & Campbell, M. A. (1989) *J. Immunol.* **143,** 2961–2968.
12. Hohn, H. A., Popescu, N. C., Hanson, R. D., Salvesen, G. & Ley, T. J. (1989) *J. Biol. Chem.* **264,** 13412–13419.
13. Sha, W. C., Nelson, C. A., Newberry, R. D., Kranz, D. M., Russell, J. H. & Loh, D. Y. (1988) *Nature (London)* **335,** 271–274.
14. Carle, G. F. & Olson, M. W. (1987) *Methods Enzymol.* **155,** 468–482.
15. Brownstein, B. H., Silverman, G. A., Little, R. D., Burke, D. T., Korsmeyer, S. J., Schlessinger, D. & Olson, M. W. (1989) *Science* **244,** 1348–1351.
16. Popescu, N. C., Amsbaugh, S. C., Swan, D. C. & DiPaolo, J. (1985) *Cytogenet. Cell Genet.* **39,** 73–74.
17. Schmid, J. & Weissman, C. (1987) *J. Immunol.* **139,** 250–256.
18. Trapani, J. A., Klein, J. L., White, P. C. & Dupont, B. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 6924–6928.
19. Masson, D. & Tschopp, J. (1987) *Cell* **49,** 679–685.
20. Brunet, J. F., Dosseto, M., Denizot, F., Mattéi, M. G., Clark, W. R., Haqqi, T. M., Ferrier, P., Nabholz, M., Schmitt-Verhulst, A. M., Luciani, M. F. & Golstein, P. (1986) *Nature (London)* **322,** 268–271.
21. Tucker, K. A., Lilly, M. B., Hech, L., Jr., & Rado, T. (1987) *Blood* **70,** 372–378.
22. Ponte, P., Ng, S. Y., Engel, J., Gunning, P. & Kedes, L. (1984) *Nucleic Acids Res.* **12,** 1687–1696.
23. Klein, J. L., Shows, T. B., Dupont, B. & Trapani, J. A. (1989) *Genomics* **5,** 110–117.
24. Harper, K., Mattéi, M. G., Simon, D., Suzan, M., Guénet, J. L., Haddad, P., Sasportes, M. & Golstein, P. (1988) *Immunogenetics* **28,** 439–444.
25. Croce, C. M., Isobe, M., Palumbo, A., Puck, J., Ming, J., Tweardy, D., Erikson, J., Davis, M. & Rovera, G. (1985) *Science* **227,** 1044–1047.
26. Hockett, R. D., de Villartay, J. P., Pollock, K., Poplack, D. G., Cohen, D. I. & Korsmeyer, S. J. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 9694–9698.
27. Welgus, H. G., Connolly, N. L. & Senior, R. M. (1986) *J. Clin. Invest.* **77,** 1675–1681.
28. Davis, M. M. & Bjorkman, P. J. (1988) *Nature (London)* **334,** 395–402.
29. Satyanarayana, K., Hata, S., Devlin, P., Roncarolo, M. G., DeVries, J. E., Spits, H., Strominger, J. L. & Krangel, M. S. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 8166–8170.
30. Forrester, W. C., Thompson, C., Elder, J. T. & Groudine, M. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 1359–1363.
31. Tuan, D., Solomon, W., Li, Q. & London, I. M. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 6384–6388.
32. Grosveld, F., van Assendelft, G. B., Greaves, D. R. & Kollias, G. (1987) *Cell* **51,** 975–985.
33. van Assendelft, G. B., Hanscombe, O., Grosveld, F. & Greaves, D. R. (1989) *Cell* **56,** 969–977.
34. Forrester, W. C., Novak, U., Gelinas, R. & Groudine, M. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 5439–5443.
35. Griesser, H., Tkachuk, D., Reis, M. D. & Mak, T. W. (1989) *Blood* **73,** 1402–1415.
36. McGuire, E. A., Hockett, R. D., Pollock, K. M., Bartholdi, M. F., O'Brien, S. J. & Korsmeyer, S. J. (1989) *Mol. Cell. Biol.* **9,** 2124–2132.
37. Kagan, J., Finger, L. R., Letofsky, J., Finan, J., Nowell, P. C. & Croce, C. M. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 4161–4165.
38. Bogenberger, J. & Haas, M. (1988) *Oncol. Res.* **3,** 301–312.